MEDICAL

HUMAN BIOLOGY

a record of research

SEPTEMBER, 1947

Vol. 19 No. 3

Published

THE JOHNS HOPKINS PRESS

HUMAN BIOLOGY

a record of research

FOUNDED BY RAYMOND PEARL

Edited by

CHARLES P. WIMSOR

in association with

A. W. FREEMAN A. D. LANGMUIR ASHLEY MONTAGU
CURT P. RICHTER ADOLPH H. SCHULTZ

Managing Editor

FRANCES D. HOSKINS

ADVISORY BOARD

ED. CLAPARÈDE
University of Geneva

S. R. DETWILER
Columbia University

EUGEN FISCHER
Kaiser Wilhelm Institut für Anthropologie, Berlin-Dahlem

CORRADO GINI
R. Università de Roma

MAJOR GREENWOOD
School of Hygiene and Tropical Medicine, London

SHINKISHI HATAI
Tôhoku Imperial University, Sendai

HANS KNIPSCHEER
Universitäts-Nervenklinik, Marburg

A. J. LOTKA
Metropolitan Life Insurance Company

D. PEYRONY
Musée Préhistorique, Les Eyzies (Dordogne)

P. RIVET
Universidad de Bogotá, Colombia

JOHN B. WATSON
New York

CLARK WISSLER
American Museum of Natural History

HUMAN BIOLOGY is published quarterly by The Johns Hopkins Press, Baltimore, Maryland, U. S. A. The advance subscription price is $5.00 a year (foreign postage 35 cents; Canadian 15 cents). Single current issues are $2.00 each; back numbers, $3.00 each; back volumes, $6.00; plus postage. Subscriptions for less than a year are charged at the single issue price.

Subscribers should notify the publishers of change in address at least two weeks before publication of the issue with which change is to take effect. No claim for non-receipt of an issue will be entertained unless made immediately upon receipt of the next succeeding issue.

Manuscripts and communications regarding editorial matters should be addressed to Editor, Human Biology, 615 N. Wolfe St., Baltimore—5, Maryland, U. S. A.

THE JOHNS HOPKINS PRESS, *Publishers*
BALTIMORE—18, MARYLAND, U. S. A.

Entered as Second Class matter, February 18, 1929, at the Post Office at Baltimore, Md., under the Act of March 3, 1879.

Copyright 1949 by The Johns Hopkins Press

HUMAN BIOLOGY
a record of research

SEPTEMBER, 1947

VOL. 19 No. 3

FAMILY STUDIES IN THE EASTERN HEALTH DISTRICT

IV. PERMANENCE OF RESIDENCE WITH RESPECT TO VARIOUS FAMILY CHARACTERISTICS*

BY H. M. C. LUYKX, D. Sc.

I. INTRODUCTION

THE purpose of this study is to delineate some of the characteristics of an urban population which are not usually obtainable from data such as those accumulated by the decennial Federal census. A wealth of material has been gathered concerning the Eastern Health District of Baltimore by various studies and surveys. The records made for three special censuses, in 1933, 1936, and 1939, have been assembled in such a way as to lend themselves particularly to analyses using the family as a basic unit. For each family surveyed it has been possible, with considerable accuracy, to determine whether it was present at a preceding or succeeding survey. Coupled with various sociological characteristics, such as nativity of

* From the Department of Biostatistics (Paper No. 235), School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, Md.

The writer is deeply grateful to the Commonwealth Fund for a fellowship, and to the College of Medicine of New York University for a leave of absence, without both of which the opportunity to make the present study could not have been realized.

the family head, size of the family, etc., it is therefore possible to present a dynamic picture of this population on a family basis. An attempt is made here to show how certain of these characteristics influence the stability of a family, as portrayed by the ability to identify the same family at a later survey.

The Eastern Health District

The Eastern Health District and its population have been described in some detail in earlier studies by Valaoras (7) in 1936, and by Reed, Fales and Badger (5) in 1942.

In a general sense, the District contains a population resembling that of the whole city reasonably well. It comprises chiefly a low- and middle-income residential area. As an administrative unit of the Baltimore City Health Department the Eastern Health District was enlarged in 1938. In order to maintain comparability, however, the present study relates only to that portion lying within the boundaries of the old District, shown in Fig. 1.

The proportion of colored persons is almost identical with that for the city as a whole. With respect to age, the white and colored populations both are slightly younger than those of the entire city. The percentage foreign-born is slightly higher, although for both the District and the city this percentage is decreasing, reflecting a low rate of immigration from foreign countries in recent years. The socio-economic (occupational) distributions of the population, for the District and for the city, also correspond fairly closely. Although it is not to be implied that data concerning the Eastern Health District may be substituted for a picture of Baltimore, it is clear that we are not dealing with an exceptional or unusual urban population.

Accumulation of the data

This study is based on three special censuses, conducted in 1933, 1936, and 1939 respectively. The first and last of these were planned and supervised by the Department of Biostatistics at the School of Hygiene and Public Health. The second survey was part of a nationwide study conducted by the U. S. Public Health Service (3), as a W. P. A. project, the supervision for the Eastern Health District, however, also being by persons associated with the Department of Biostatis-

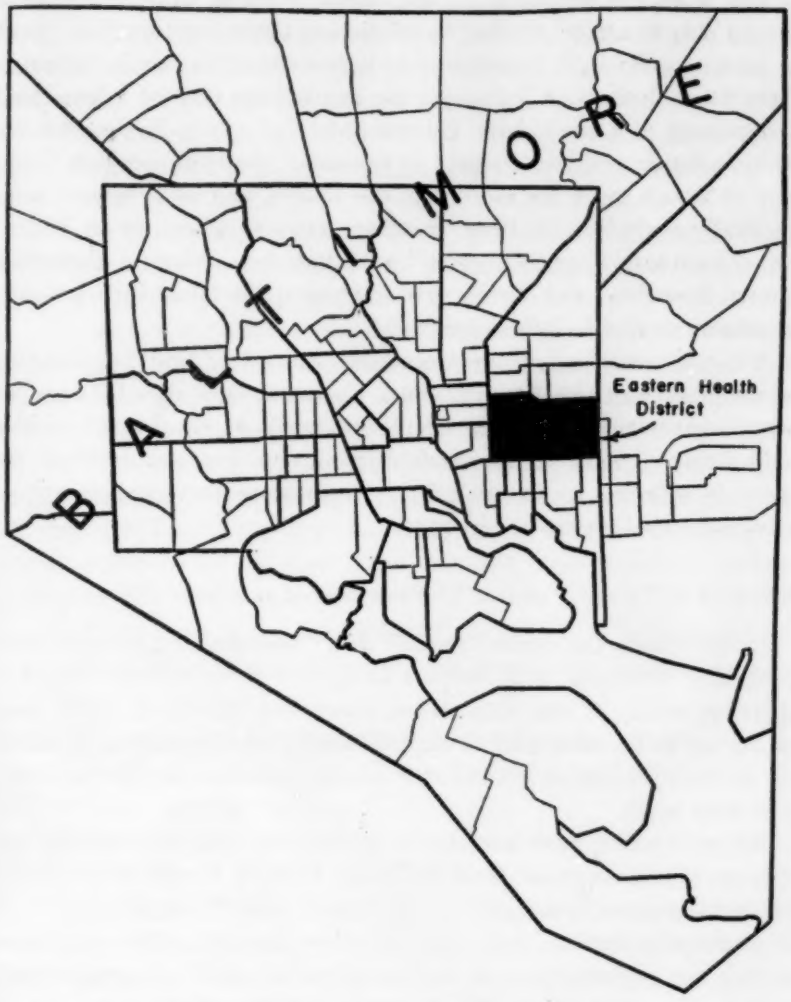

FIG. 1. MAP OF BALTIMORE, SHOWING THE EASTERN HEALTH DISTRICT

tics of the School of Hygiene. The general scheme in each census was to send an enumerator to a list of addresses—all the houses in one block, let us say—where he would fill out a questionnaire for each household by interviewing a member of the family who was at home at the time.

The written instructions to the enumerators in the three studies differed only in minor details. In relation to the present analysis, dealing with households, it is pertinent to note some difference in definition of the term "household." In 1933 the instructions defined a household as consisting of a head, wife, children, and "all others living with the family, whether relatives, lodgers or servants." In 1936 members living away at school, away for more than one month, and all "visitors" were specifically excluded. In 1939 "members away at school, in an institution, reform school, or CCC camp" were included. Where a distinction between household and family was necessary, the latter included only the related members of the household.

After accumulation of the schedules, careful editing for accuracy and consistency of the data was done. Correlations of data between the several censuses have further established levels of consistency. For a study throwing light on the variability of data between censuses the reader is referred to an analysis of consistency in reporting age of individuals, by Densen (2) in 1940.

Definition of "family" and of "permanence of residence"

In this study the terms "family" and "household" are used interchangeably, referring to all persons living at a given address, except as otherwise noted. It has been shown above that definitions differ from one survey to the next only as regards family members away at school, etc. In this population the number of such persons may be presumed to be very small.

Several studies have been made in the past with reference to permanence of residence of families in the Eastern Health District. On occasion the terms "immigration" and "emigration" have been used, on the assumption that a family present at one census and not present at another has migrated into or out of the study area. Several circumstances can, and undoubtedly do, occur which make the use of the term "migration" inexact in the strictest sense. There are families which disappear by reason of change of name, by marriage perhaps, or the death of the individual who has been the only means of identifying the family. Families may be created by a member of a former family breaking away and setting up an establishment of his own. These events have not been recorded in such a way that they could be counted separately for this analysis. Furthermore, there is evidence that actual migra-

tion has occurred where a family has moved out of the study area and back, or vice versa, between two censuses. A loose definition of immigration and emigration, therefore, includes some cases which are not strictly migration, and excludes some cases which are. There is reason to believe that variations caused by these factors alone are not large, even though they must not be ignored in any discussion of permanence of residence.

The term "permanence of residence" is used in this paper as coming the closest to a proper description of all that is implied, and is borrowed from the study by Valaoras (7).[1]

An important factor is the criterion by which a family at one census is identified as the same family at another census. In this study, permanence of residence is based on whether a family enumerated at one census, was or was not present in the Eastern Health District at the next census. Some confusion might be caused by the fact that a family may undergo a definite change in composition, and that from one point of view the resulting unit is the same family as the original one, and from another point of view it is not. We have therefore defined a family as still in existence if

(1) the head of the family is still the head at a subsequent census (where husband and wife are both present the husband is always listed as the head), or

(2) a wife has become a head at a subsequent census.

Identification on a more complex basis has not proved practical in dealing with the existing records.

Table 1 shows the number of families counted in the present study, and the number of original families[2] remaining at the next census in each instance. The effect of varying criteria for identification of families at a subsequent census may be observed by comparing these data with figures for "emigration" presented in other analyses. Badger (1) studied the same households used in this paper, but identified a family as still present at a subsequent survey if the head was still present *in any capacity,* i. e., he or she might be an in-law of the head of a new

[1] Valaoras defined a "settled" family as one which was identified as having existed as the same family at a previous study eleven years earlier (1933 and 1922).

[2] Throughout this study the term "original families" refers to those families counted in the first of two succeeding censuses.

TABLE 1

Number of families in the Eastern Health District, 1933, 1936 and 1939, by color

	TOTAL		WHITE	COLORED
	Number	Per cent		
Total families surveyed in 1933	13 710	100.0	10 839	2 871
Heads still heads in 1936	*9 500*		*7 835*	*1 665*
Wives who became heads by 1936	*354*		*267*	*87*
Total families remaining in 1936	9 854	71.9	8 102	1 752
Heads still heads in 1939	*7 334*		*6 169*	*1 165*
*Wives who became heads by 1939**	*271*		*210**	*61**
Total 1933 families remaining in 1939	7 605	55.5	6 379	1 226
Total families surveyed in 1936	13 954	100.0	10 928	3 026
Heads still heads in 1939	*9 207*		*7 512*	*1 695*
Wives who became heads by 1939	*435*		*337*	*98*
Total 1936 families remaining in 1939	9 642	69.1	7 849	1 793

* It was not possible to count the wives who became heads between 1936 and 1939 separately. The numbers are estimated on the assumption that the ratio of wives who became heads in the second interval to those in the first interval was the same as the ratio of heads who remained heads in the second interval to those in the first interval:

$$267 \times \frac{6169}{7835} = 210, \qquad 87 \times \frac{1165}{1665} = 61.$$

family which by this definition would then be called the same family. Phair (4) was concerned with emigration in a study of tuberculosis in the Eastern Health District. He identified families by name at a subsequent census, and when there was doubt as to whether a household should be called the same or not, the decision was based on "a concensus of opinion" of persons familiar with the problem.

We see that the present study employs a stricter definition than either Badger or Phair did, with the result that a smaller proportion of families is identified as still present than was the case for the other two studies. Phair has calculated an annual rate of "emigration" for the period 1933 to 1936. Comparable rates have been calculated for the data from Badger and from the present study, and are shown in Table 2, which shows how the annual emigration rates become higher

TABLE 2

Emigration rates based on various methods of identifying families

COLOR AND BASIS OF FAMILY IDENTIFICATION	ANNUAL EMIGRATION RATE (PER CENT)		
	1933 to 1936 (2.5 yrs.)	1936 to 1939 (3.5 yrs.)	1933 to 1939 (6.0 yrs.)
White:			
(1) Former head is present in any capacity (Badger)	8.0		6.1
(2) Name and "consensus of opinion" (Phair)	8.8		
(3) Head remains head or wife becomes head (present study)	10.1	8.0	6.9
Colored:			
(1) Former head is present in any capacity (Badger)	12.1		8.2
(2) Name and "consensus of opinion" (Phair)	14.3		
(3) Head remains head or wife becomes head (present study)	15.6	11.6	9.6

as the rules for identifying a family at a subsequent census become more limiting (lines downward). The emigration rates for colored are much higher than for whites, but follow the same pattern, — both as to the effect due to different methods of identifying families at succeeding surveys and the length of the interval involved (see below).

Rider and Badger (6) have shown that the rate of emigration decreases with the length of time a family has been established. Table 2 shows this effect clearly also, the annual rate of loss of families becoming smaller as the interval covered by the analysis is increased (columns from left to right). A more convenient way of stating this

is to calculate the proportion of families remaining at the end of a stated interval, out of those counted originally. There is good evidence (which space does not permit us to demonstrate here) that for any given group of families, if the percentage remaining out of the original total were plotted on a time scale, the curve sloping downward to the right, would fall more rapidly than an exponential at first and would flatten out, falling less rapidly than an exponential, later on. If this curve were an exact exponential, we would say that the proportion of families lost per unit time interval (i. e., the probability of a family's leaving the survey) was a constant. Hence the evidence is that this is not so, which is in agreement with the findings of Rider and Badger (6) cited above.

The following sections are devoted to demonstrating the associations which exist between the proportions of families remaining at the end of the stated interval and certain family characteristics. Of the various characteristics discussed, ownership of the home seems to have the most striking correlation with "stability" of the family. Differences in proportion remaining which appear to be associated with color, for example, disappear when each race is distributed by ownership of the home. The first analysis therefore deals with home ownership, and subsequent topics are made specific for ownership.

In most instances two observations of proportions remaining have been made, for the early and for the late census interval, i. e., proportion of 1933 families remaining in 1936, and proportion of total 1936 families remaining in 1939. Where this is true, the total 1933 and 1936 original families have been used as the basis of the proportion, and the number remaining in 1936 and 1939 out of their respective original groups, as the numerator of the proportion. The effect is to obtain a larger sample and possibly compensate for sampling errors and secular trends. The average interval is then taken as three years. In two instances, dealing with nativity and education, only the second, 3.5-year interval was observed, and in one case, monthly rental, only the first, 2.5-year interval was used.

Family characteristics

Some of the characteristics under discussion are permanent in the sense that a given family does not normally change from one class to another. Nativity and education of the family head will not change.

Age of the head is the same sort of characteristic, except that there is a fixed increase for all family heads equal to the study interval. There are exceptions, of course, in those cases where the family head changes. It has been pointed out that these were relatively few.

The other characteristics studied in the following sections are such that a family may more readily change from one class to another. A home owner may become a renter three years later, or vice versa. Family type may change from "with children" to "without children." Size of family, number of wage-earners and socio-economic class are all characteristics subject to change for a given family. In some instances the passage of time tends to produce changes in a specific direction, such as increasing size of family, particularly where the family head is young. In addition there are instances of families changing from one category to another fortuitously, such as by reason of inaccurate enumeration. Densen (2) has demonstrated the discrepancies which may exist between reported ages of the same individual at different censuses. In the following discussions families are classified according to their characteristics as originally reported, in 1933 or 1936. The families were identified at the end of the interval in question mainly by name of the head or his wife, without reference to any categorical classification. Thus the proportion of home owners remaining, for example, means the proportion who were originally home owners, even though at the end of three years some of them may have become home renters.

The manner of calculating proportions remaining should be kept in mind throughout the following sections. In each case, families have been classified into sub-groups, and the proportion remaining is the ratio of the number remaining to the number originally counted, for each sub-group separately. In this way we know that the denominator of each proportion includes all the families comprised in the numerator.

Throughout this study, data for white and colored families have been kept separate on the basis that the obvious racial, social and economic differences render these two groups essentially non-comparable.

II. PROPORTIONS OF FAMILIES REMAINING AFTER A FIXED AVERAGE TIME INTERVAL, WITH RESPECT TO VARIOUS FAMILY CHARACTERISTICS

Ownership of the home and rental

Table 3 shows the percentage of families owning their homes in

TABLE 3

Distribution of families and proportions remaining at the end of 3 years, by color and home ownership (1933 + 1936)

	WHITE	COLORED
Total number of original families (1933 + 1936)*	21,705	5,889
Proportion remaining	.733	.601
Number owning their homes	13,246	431
Proportion remaining	.833	.807
Number renting their homes	8,459	5,458
Proportion remaining	.576	.585
Percentage owning their homes	61.0	7.3

* "Unknown" as to ownership are excluded.

each color group, this being 61 for whites and 7 for colored. The proportion remaining for whites is .733, and for colored .601, for an average 3-year interval, which is a highly significant difference. Table 3 also shows that for home owners the proportions remaining for white and colored are .833 and .807 respectively ,and for home renters the respective proportions remaining are .576 and .585. These differences are entirely without significance. We may suspect, therefore, that the large difference between proportions remaining for colored (.601) and white (.733) is due to the fact that such a small percentage of the colored own their homes.

Figs. 2 and 3 show the distribution of families by monthly rental paid, and the proportion remaining for each rental sub-group. It is not surprising to note that the rental paid by whites is distributed over a

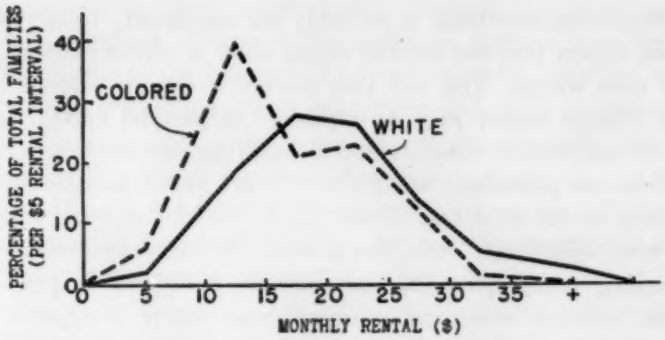

FIG. 2. DISTRIBUTION OF WHITE AND COLORED FAMILIES BY RENTAL (HOME RENTERS, 1933 ONLY) (See Appendix, Table 8.)

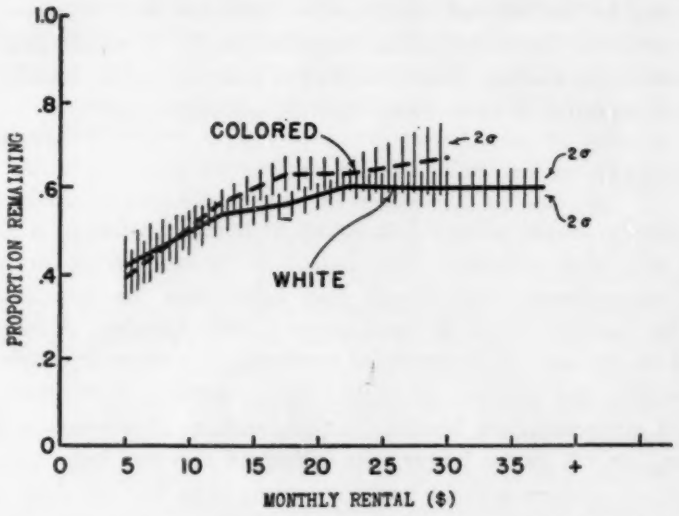

FIG. 3. PROPORTIONS OF FAMILIES REMAINING AT THE END OF 2.5 YEARS, BY COLOR AND RENTAL (HOME RENTERS, 1933 ONLY) (See Appendix, Table 8.)

The vertical shading of these curves is intended to show roughly the extent to which they might vary due to sampling.

higher part of the scale than that of the colored. The curves for proportion remaining show clearly that those paying a higher rent are the more permanent residents. Inspection of the graph indicates that for any particular rental class the difference between the colored and

white proportions remaining is probably not significant. Intuitively one might not expect that the colored would show a *higher* proportion remaining than whites. The fact that the curve for the colored is consistently slightly higher may be explained in part by several factors. Nearly all the colored families are included among the home renters, whereas the less permanent white residents are in this class—only about 40 per cent of the total (see Table 3). If the difference is real, even though not "significant," this would deny the suspicion mentioned in the preceding paragraph. The fact that the difference in proportions remaining between white and colored home renters disappears might then be due in part also to the fact that the colored are in a lower rental bracket than the whites, and the greater stability of colored home renters is offset by the lesser stability of the low-rental group. Another factor may be that the colored are more restricted in their movements, and hence compelled to remain in one place longer. Also, the residential character of the Eastern Health District is such that white families have a greater incentive to move away than the colored.

Family type

A family, as the picture first comes to mind, consists of a married couple with their children. This indeed is the most common type of family encountered, even though that description fits less than half of all the families found in the Eastern Health District. A large proportion of the families enumerated consisted of a married couple without children, and another, similarly large proportion were those which included other members besides children, such as grandparents, grandchildren, etc. A group larger than either of the two last named was made up of women without husbands, more than half of these having children, grandchildren, or other relatives living with them.

How to classify families as to type depends on the purpose of such classification. From a biologic point of view, direct blood relationship is the chief factor to consider. From an economic point of view, the number of mouths to be fed is the important thing. The total number of combinations of the type of relationships which may exist in a family, differentiating between specific relationships such as grandparents, grandchildren, in-laws, etc., has provided a set of 192 classes of family type. It is obviously not practical to maintain such an extended classification, and it would serve no purpose here. Permanence of residence

would seem to be related chiefly to the economic and social aspects of the family structure. In this study, therefore, families have been classified with respect to the presence of (a) a spouse, (b) children, and with respect to the sex of the family head when no spouse is present.

In order to avoid the task of referring to the original family records to determine the age of the individual family members, families with children are here defined as all households in which children and/or grandchildren of the family head are present. Males and females with no spouse present may be unmarried, or they may be widowed, divorced or separated. In the group with children, the latter is most likely the case. The single persons who are bachelors are chiefly without children. Here it must be remembered that for a household of bachelors, one of them was listed as the "head of the household" and the others as non-related members. In the population, therefore, the number of males and females with no spouse present is greater than the number of families with non-married heads without children. The number of *families* consisting of a single member, discussed under family size in a later section, gives the number of lone men and women who are listed as family heads (whether living with others or not). The number of persons living alone is given by the number of *households* of size one.

The relative proportions remaining at the end of 3 years will be considered for family type groups based on these relationships and indicated in Table 4. This table shows the proportions remaining after 3 years, for each family type, as well as the percentage owning their homes. The great difference between percentage of white families owning their homes, and the percentage of colored, has been pointed out. The families with children have larger proportions remaining than those without,—and in both of these groups the married family heads show the greatest permanence of residence. An unexpected observation is that among the whites the family heads with no spouse present include the largest percentages owning their homes, and at the same time, specific for ownership as well as for color, these are generally the classes which have the lowest proportions remaining.

It is also significant that, of the family heads with no spouse present, much the greater number are female. That these show higher proportions remaining than the corresponding males is somewhat unexpected. By our definitions, single male family heads, if they married, would subsequently still be identified as the same family heads. But

TABLE 4

Distribution of families and proportions remaining at the end of 3 years, by color, home ownership and family type (1933 + 1936)

COLOR AND FAMILY TYPE	PERCENTAGE OWNING THEIR HOMES	HOME OWNED			HOME RENTED		
		Original number of families	Proportion remaining	σ	Original number of families	Proportion remaining	σ
White:							
Families with children: Total	62.8	10 054	.860	.004	5 959	.620	.006
Married heads	62.2	7 894	.901	.003	4 800	.650	.007
Heads with no spouse present:							
Male	77.2	439	.647	.023	130	.438	.044
Female	62.6	1 721	.729	.011	1 029	.500	.016
Families without children: Total	56.1	3 192	.747	.008	2 500	.471	.010
Married heads	52.8	1 943	.836	.008	1 738	.521	.012
Heads with no spouse present:							
Male	61.6	406	.522	.025	253	.308	.029
Female	62.4	843	.650	.016	509	.381	.022
Colored:							
Families with children: Total	6.6	235	.826	.025	3 315	.641	.008
Married heads	7.4	182	.890	.023	2 287	.684	.010
Heads with no spouse present:							
Male	6.7	8	.750	.153	112	.438	.047
Female	4.7	45	.578	.074	916	.559	.016
Families without children: Total	8.4	196	.786	.029	2 143	.498	.011
Married heads	9.8	142	.859	.029	1 300	.571	.014
Heads with no spouse present:							
Male	4.2	16	.562	.124	369	.363	.025
Female	7.4	38	.605	.079	474	.403	.022

single female family heads would no longer be family heads after they married, and that former family would then be lost from the study. If this condition was sufficiently frequent to introduce appreciable differences between male and female proportions remaining, then without it the differences between male and female would be even greater than shown in the table.

In this particular discussion, proportions remaining are determined mainly by emigration, and it is difficult to put one's finger on the proper explanation for any associations which may exist between emigration and family type. A study of changes in family type over a period of years, not attempted here, would perhaps throw light on this question.

Socio-economic class

Families are classified here according to socio-economic class of the oldest worker in the family. In nearly 90 per cent of the cases this was the head of the family, and in the others some member related to the head. The occupation of boarders or lodgers was never used to determine socio-economic classification, since *family* is here defined in the restricted sense (no non-relatives included). Families in the servant class are only those not living in the homes of their employers. A frequency distribution of families by socio-economic class is shown in Fig. 4.

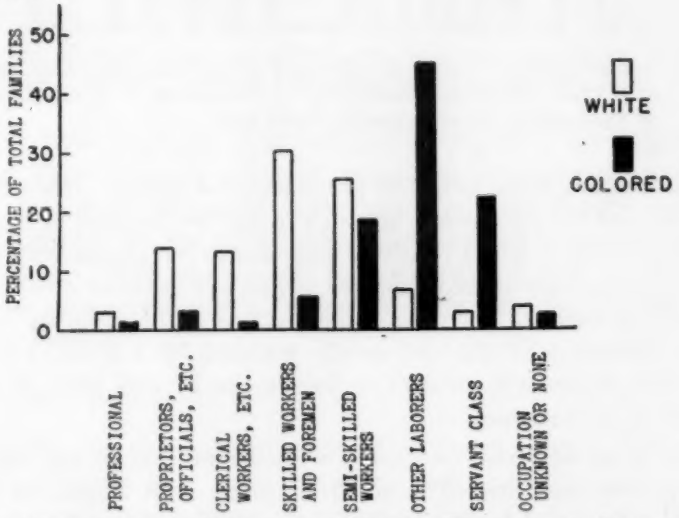

FIG. 4. DISTRIBUTION OF WHITE AND COLORED FAMILIES BY SOCIO-ECONOMIC CLASS (1933 + 1936) (See Appendix, Table 9.)

Figure 5 A shows the proportions of white families remaining for each socio-economic class arranged in order of frequency of that class. Figure 5 B is the same thing for colored families. The order of relative frequency for the classes is quite different between the white and colored, the largest group for whites being skilled workers and foremen, and for colored, ordinary laborers. Figure 5 A shows a tendency

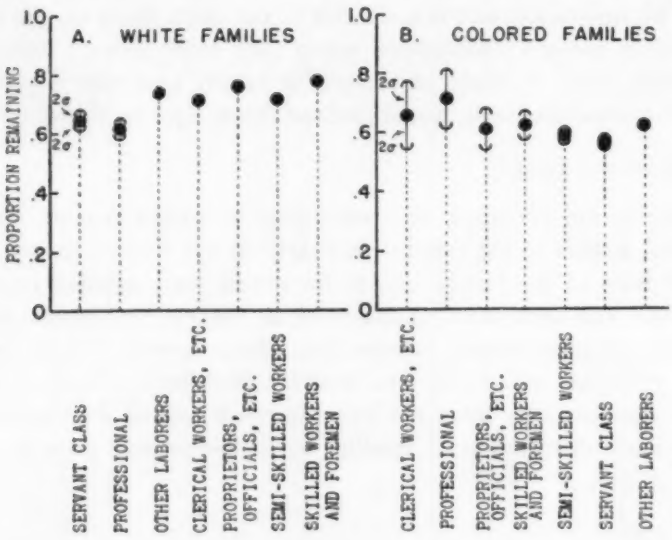

FIG. 5. PROPORTIONS OF WHITE AND COLORED FAMILIES REMAINING AT THE END OF 3 YEARS, BY SOCIO-ECONOMIC CLASS (ARRANGED IN ASCENDING ORDER OF FREQUENCY) (See Appendix, Table 10.)

for the smaller groups to be less permanent, for whites. This tendency is not evident for the colored, although it may be obscured by variation due to sampling, which is relatively large due to the groups being small. There is some logic in expecting the largest occupational groups to be the most permanent. If, for example, the type of home in the Eastern Health District is of the sort usually occupied by a laborer's family, then such professional families as live in the District have an added incentive to move away.

The effect of home ownership is shown in Fig. 6, and although much greater than the effect of classification with respect to socio-economic class, it does not seem to affect the relative stability of the

different classes. (This breakdown is shown for whites only, because so few of the colored families are home owners. The proportions remaining for colored home renters are almost identical with those for all colored families.) When occupational groups are placed in order of permanence for home owners (as in this figure), this is also the order of permanence for home renters. So it may be said that dif-

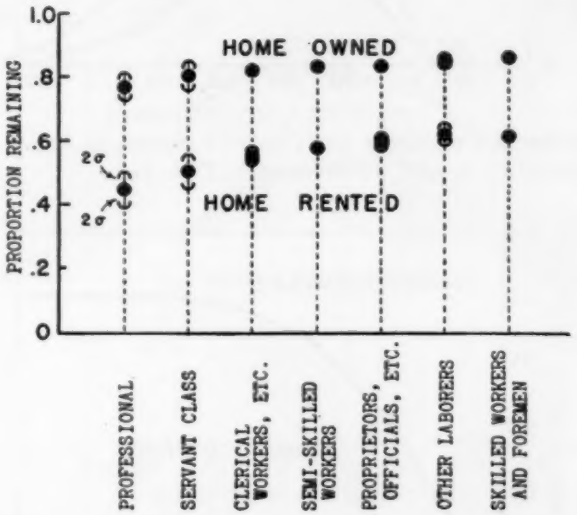

FIG. 6. PROPORTIONS OF WHITE FAMILIES REMAINING AT THE END OF 3 YEARS, BY OWNERSHIP OF THE HOME AND SOCIO-ECONOMIC CLASS (ARRANGED IN ASCENDING ORDER OF PERMANENCE) (See Appendix, Table 11.)

ferences between groups in percentages owning their homes (see Appendix, Table 10), are not great enough to materially affect the relative proportions remaining, between groups, even though the order of frequency is not the same for home owners and home renters. For home owners, skilled workers and foremen are the largest group, whereas for home renters, semi-skilled workers are the most frequent (Appendix, Table 11).

Age of the family head

Of the various family characteristics (besides home ownership), the age of family head is the one which perhaps has the most striking association with permanence of residence. Figs. 7 and 8 show the fre-

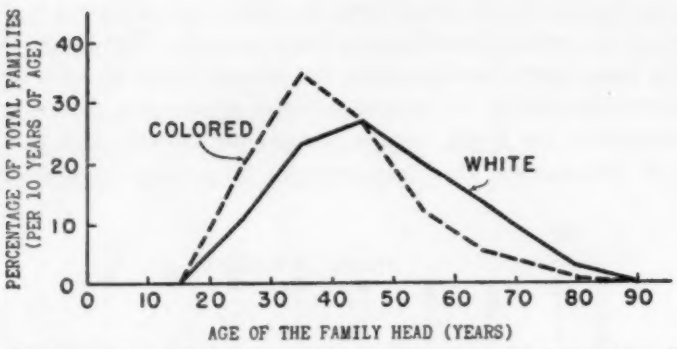

FIG. 7. DISTRIBUTION OF WHITE AND COLORED FAMILIES, BY AGE OF THE FAMILY HEAD (1933 + 1936) (See Appendix, Table 12.)

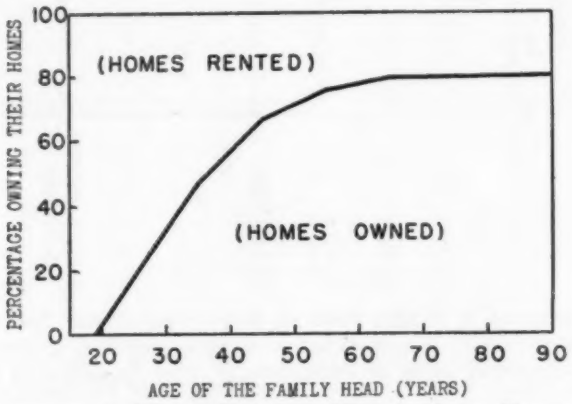

FIG. 8. PERCENTAGE OF WHITE FAMILIES OWNING THEIR HOMES, BY AGE OF THE FAMILY HEAD (1933 + 1936) (See Appendix, Table 12.)

quency distribution and percentage of whites owning their homes, by age of the head. The colored family heads are much younger than the whites. The age of head associated with maximum stability is greater for colored than for white home owners, as shown in Fig. 9. Maximum permanence of residence for white home owners is around age 45. For white home renters and for colored it is closer to age 55. As pointed out earlier, the colored are mostly home renters.

The smoothness of the curves of proportion remaining by age of the head is evidence of the strong association between these factors. At the younger ages the percentage of home owners is naturally very

low, as is seen in Fig. 8. But even for home owners only, the permanence of residence increases with age up to a certain point. Beyond the age of maximum proportion remaining, the stability of families decreases as age of the head increases. This drop probably represents families breaking up or otherwise losing their identity by reason of death of the head. In other words, the low proportions remaining at the young

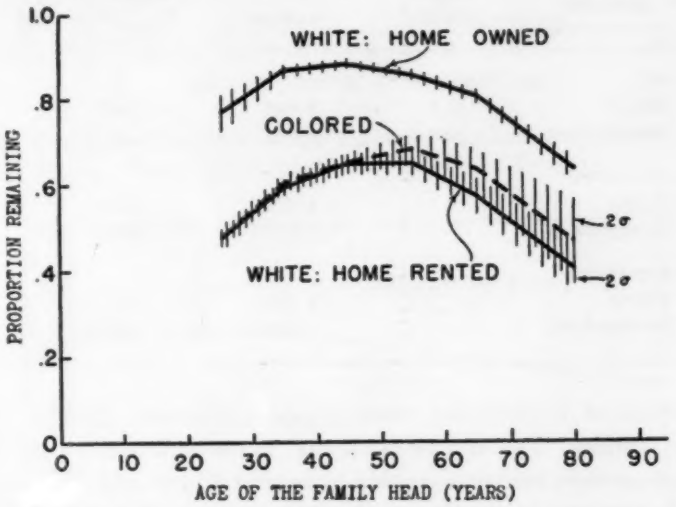

FIG. 9. PROPORTIONS OF FAMILIES REMAINING AT THE END OF 3 YEARS, BY COLOR, OWNERSHIP OF THE HOME AND AGE OF THE FAMILY HEAD (See Appendix, Table 12.)

ages are probably caused chiefly by migration, whereas at the old ages they are probably due to mortality. In this particular, the fact that the average interval of observation is three years is of importance. If the interval were longer, not only would the entire proportions-remaining curve be lower, but the effect of death would reach into the lower ages, thus probably shifting the peak of the curve to the left.

Nativity of the family head

Table 5 shows the proportions of families remaining by nativity of the family head. The data are available for the later survey interval only, and since there are practically no foreign-born colored in the Eastern Health District, this comparison is made for whites only. The

TABLE 5

Proportions of white families remaining at the end of 3.5 years, by home ownership and nativity (1936 only)

HOME OWNERSHIP AND NATIVITY	PERCENTAGE OWNING THEIR HOMES	ORIGINAL NUMBER OF FAMILIES	PROPORTION REMAINING	σ
Total:				
Native	55.7	8 105	.708	.005
Foreign-born	70.2	2 821	.749	.008
Home owned:				
Native		4 504	.820	.006
Foreign-born		1 976	.818	.009
Home rented:				
Native		3 581	.565	.008
Foreign-born		840	.588	.017

proportion of foreign-born remaining is significantly greater than for native families. This can be attributed to the fact that 70 per cent of the foreign-born are home owners, as against 56 per cent for the natives. This in turn can be ascribed to the fact that immigration to the United States has declined sharply in the years prior to 1936, and the entire foreign-born population is therefore older than the native. When separated by ownership of the home, there is no significant difference between native and foreign-born proportions remaining.

Education of the family head

Figs. 10 through 12 show the distributions and proportions remaining by education of the family head. In Fig. 10 we see that the average education for whites is higher than for colored. For both races more family heads received no schooling at all than went to school as far as first or second grade only. If they went to school, most of the whites finished grammar school, whereas most of the colored went no farther than fifth or sixth grade.

Fig. 11 shows that a larger percentage of the poorly educated family heads are home owners than the better educated.

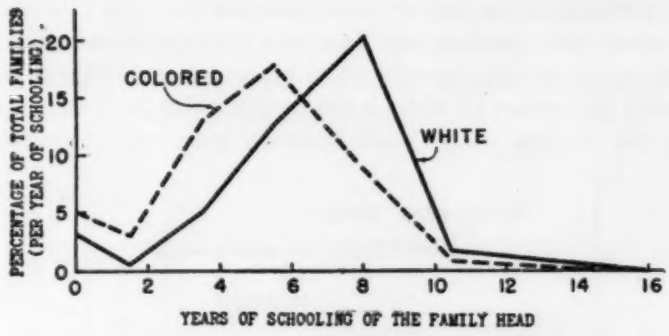

FIG. 10. DISTRIBUTION OF WHITE AND COLORED FAMILIES, BY EDUCATION OF THE FAMILY HEAD (1936 ONLY) (See Appendix, Table 13.)

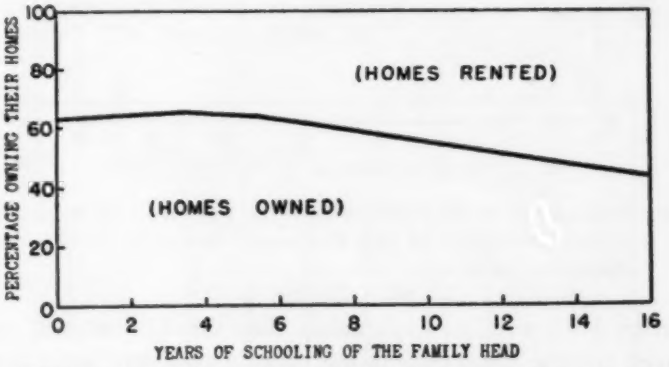

FIG. 11. PERCENTAGE OF WHITE FAMILIES OWNING THEIR HOMES, BY EDUCATION OF THE HEAD (1936 ONLY) (See Appendix, Table 13.)

Fig. 12 shows that for the white home owners permanence of residence increases slightly with educational achievement, reaching a maximum at fifth or sixth grade and beyond. This is about what would be expected. For white home renters and colored, the better educated are the less permanent residents. This can perhaps be explained by the same factor which causes the better educated to include a smaller percentage of home owners (Fig. 11), viz., the Eastern Health District is a somewhat lower than average residential section on the socio-economic scale. Residents of the District with an advanced education comprise, among others, those professional persons connected with the large hospitals and schools located there. These settle in the District only temporarily,

do not purchase homes, and are more likely to move out than those of a lower educational classification. This does not explain satisfactorily why the maximum stability among white home renters (Fig. 12), or the maximum percentage of white home owners (Fig. 11), should both be among the families whose heads have not gone beyond fourth grade.

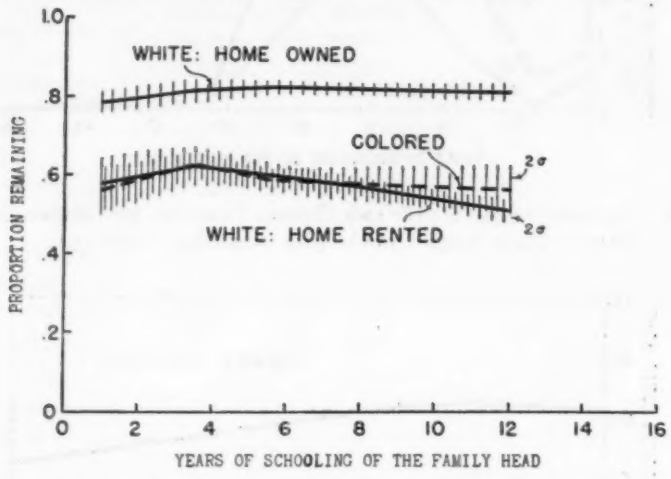

FIG. 12. PROPORTIONS OF FAMILIES REMAINING AT THE END OF 3.5 YEARS, BY COLOR, OWNERSHIP OF THE HOME AND EDUCATION OF THE HEAD (See Appendix, Table 13.)

If foreign birth were an explanation, then this phenomenon would be displayed for the home owners too, since a relatively large percentage of the foreign-born are home owners (see Table 5).

Size of family or household

The terms *family* and *household* have been used in this paper interchangeably to refer to a socio-economic unit living in one place. A family is defined as including only the related members, not omitting in-laws, step-children, adopted children, etc. A household includes in addition any servants, boarders, lodgers, or other unrelated persons living in the same home. In the great majority of cases in this study the terms are in effect synonymous, there being no unrelated persons present in the household (see Appendix, Table 14). Although the average household is inevitably slightly larger than the average family, it is not necessary to describe the relationships between proportions remaining

and both family size and household size. When substituting the word *household* for *family* in Figs. 13 through 15, the effect is merely to shift the curves to the right a fraction of one class interval (a slightly larger fraction for the colored than for the whites).

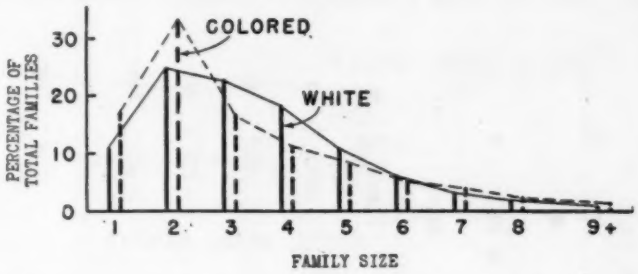

FIG. 13. DISTRIBUTION OF WHITE AND COLORED FAMILIES, BY FAMILY SIZE (1933 + 1936) (See Appendix, Table 15.)

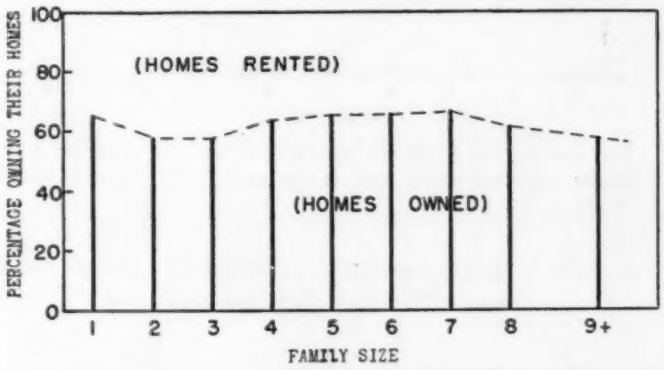

FIG. 14. PERCENTAGE OF WHITE FAMILIES OWNING THEIR HOMES, BY FAMILY SIZE (1933 + 1936) (See Appendix, Table 15.)

A comparison of percentage distributions by family size for white and colored (Fig. 13) is interesting in that it shows that the average white family is larger than the average colored family. For the whites about 64 per cent of the families have more than 2 members, whereas for the colored these constitute only about 50 per cent of all the families.

Fig. 15 shows a highly significant association between family size and proportion remaining, for white home owners as well as for home

renters (most of the colored being in the latter category). The larger families are decidedly the more stable. Where there are children present there is apparently a slightly higher tendency to own the home (Fig. 14), and a strong tendency to remain in one place. In spite of the fact that about the same percentage of families of size one as of size

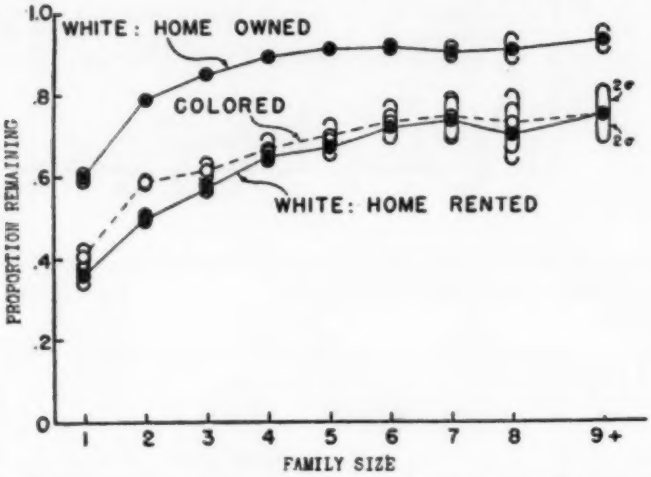

FIG. 15. PROPORTIONS OF FAMILIES REMAINING AT THE END OF 3 YEARS, BY COLOR, OWNERSHIP OF THE HOME AND FAMILY SIZE (See Appendix, Table 15.)

5 or 6 seem to be home owners, the "bachelor" families are the least likely to remain, whether they own their homes or not.

Number of wage-earners in the family

In the great majority of the cases there is one wage-earner in the family, as shown in Fig. 16, with about an equal number of families having more than one wage-earner and none. The percentage distribution of families by number of wage-earners per family is nearly identical for white and for colored.

In view of the fact that the stability of families seems to vary directly as the number of wage-earners per family, as shown in Fig. 18, it would seem that the number of wage-earners is correlated with the economic well-being of the family, regardless of home ownership. For families with one or more wage-earners this is accounted for to a slight

extent by the fact that the percentage owning their homes also increases with the number of wage-earners, as is seen in Fig. 17.

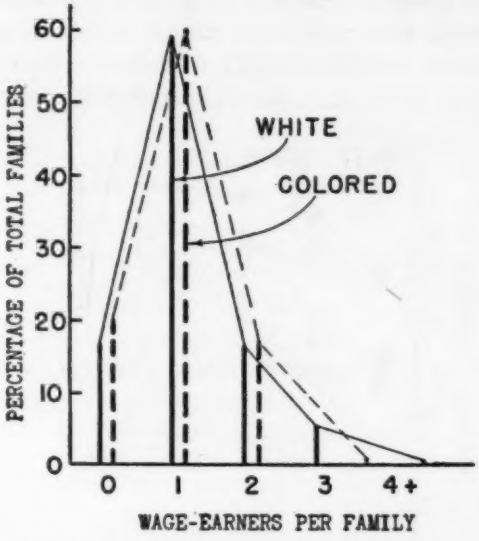

FIG. 16. DISTRIBUTION OF WHITE AND COLORED FAMILIES, BY NUMBER OF WAGE-EARNERS IN THE FAMILY (1933 + 1936) (See Appendix, Table 16.)

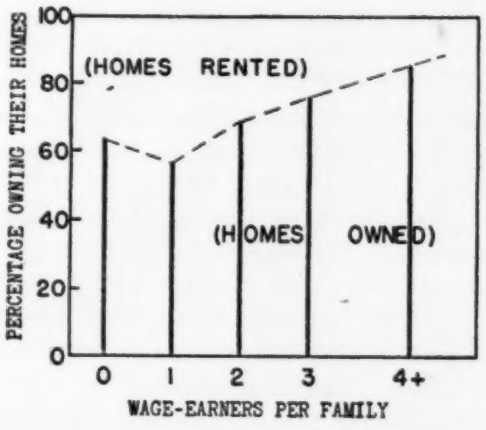

FIG. 17. PERCENTAGE OF WHITE FAMILIES OWNING THEIR HOMES, BY NUMBER OF WAGE-EARNERS IN THE FAMILY (1933 + 1936) (See Appendix, Table 16.)

For families having no wage-earners, such stability as they show seems to be due to a greater extent to the percentage owning their homes. Among the white home owners only (Fig. 18) these families are decidedly less permanent than those with one or more wage-earners. That the non-wage-earner families include a relatively high percentage of home owners may be ascribed to the fact that if the home were not

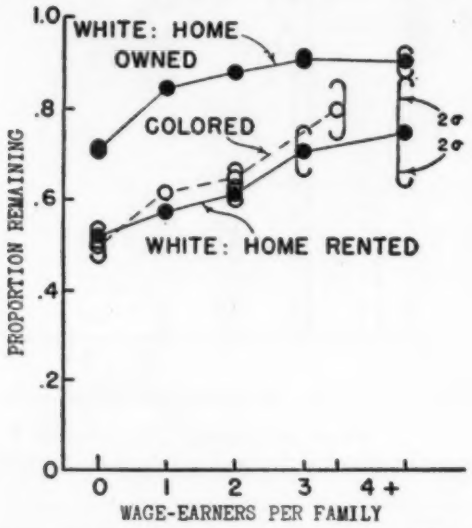

FIG. 18. PROPORTIONS OF FAMILIES REMAINING AT THE END OF 3 YEARS, BY COLOR, OWNERSHIP OF THE HOME AND NUMBER OF WAGE-EARNERS IN THE FAMILY (See Appendix, Table 16.)

owned, such families would probably break up or move away. In other words, among the home renters non-wage-earner families are relatively scarcer than among the home owners (Fig. 17).

One might suspect that more than one wage-earner in the family would be an indication of economic insecurity,—a single wage-earner not providing enough support by himself. Using the relative stability of these families, and the percentage owning their homes, as an indication, there is no evidence that this is true here. The well-to-do, where the head of the house is traditionally the only provider, are relatively scarce in the Eastern Health District. It was pointed out earlier that the groups having the largest proportion of home owners are the foreign-born and those of mediocre educational achievement. In such families, having more than one wage-earner is certainly a desirable state, and

one which enhances their economic well-being, and hence their permanence of residence.

In this connection Table 6 shows that families on relief are decidedly

TABLE 6

Proportions of families remaining at the end of 3.5 years, by color and relief status

	WHITE			COLORED		
	Original number of families	Proportion remaining	σ	Original number of families	Proportion remaining	σ
Families on relief	864	.638	.016	907	.555	.017
Non-relief families	10 008	.725	.004	2 107	.608	.011

less permanent than those not on relief.

III. SUMMARY AND DISCUSSION

The purpose of this study is to obtain a dynamic picture of what happens to families, as revealed by censuses repeated at intervals of several years. Three surveys of the Eastern Health District of Baltimore, in 1933, 1936, and 1939, have provided data for a population of some 56,000, constituting a total of a little less than 14,000 families. This population has no strikingly unusual characteristics, and resembles the population of Baltimore as a whole in that about 23 per cent are Negro, 14 per cent are foreign-born, and the age distribution resembles that of the city although the average age for the District is slightly lower. Socio-economic classification places over 50 per cent of the whites in the skilled and semi-skilled worker category, and nearly 80 per cent of the Negroes in the laborer and servant groups.

The present analysis consists of counting and classifying all the families in the District at a given census, and then determining what proportion of these families are identified as being present at a subsequent census. Identification is based on the presence of the family head, or his wife, at the later date. Families no longer present may be assumed to have moved away in the great majority of cases, as revealed

by evidence presented in other studies. Some families, however, are known to have been "lost" by reason of death of the only family head, or by being broken up for other reasons. The basic factor studied in this paper is the *proportion* of families, in particular groups or classifications, *remaining* in the District.

The present study corroborates evidence presented elsewhere, viz., that *families do not leave a study area at a constant rate, but at a decreasing rate.* That is to say, families differ in their stability, and a given group of families will decrease in size more rapidly at first, and less rapidly as the proportion of unstable families in the group decreases (i. e., the probability of a family's remaining increases). The evidence is that curves representing proportion of original families remaining, by time since original observation, are generally "sharper" than an exponential curve (the latter would be the expected shape of a curve if families were lost from the study at a uniform rate).

In comparing the characteristics associated with proportions remaining, it is evident that families in one category may be lost from the study at a greater rate than families in another category. Thus, when comparing *Negroes* with *whites,* the latter are much the more stable. Possible reasons for this difference will be reviewed presently.

Classifying families by *type,* such as the presence or absence of children, whether the head is married, and if no spouse is present, whether male or female: families with children are more stable (more "permanent") than those without; married couples are more permanent than heads with no spouse present; and of the latter the females remain in greater proportion than males.

In comparing the proportions remaining in any two groups, two reasons present themselves as being the possible cause of finding a significantly higher proportion remaining for one group than for another. First, group A may be inherently more stable than group B. In this event A may be considered a more desirable or more natural state than B. Families are not as likely to move or break up when they are in that state, and one would expect to find more families of type A than of type B at any given time. The logic of this theory is supported by several of the findings in this paper. Of the various family types, families with children are more numerous and more permanent than families without children. Likewise families headed by married couples outnumber those with single heads, and they are much more stable.

The other possible assumption is that families in given groups are disappearing and not being replaced, and that group A has a higher proportion remaining than group B because the original observation made was at a point farther along the curve of proportions remaining, i. e., a point where that curve is flattened out and no longer falling as rapidly as at first. The latter hypothesis, for example, might explain in part why the white families show higher proportions remaining than the colored. The percentage of colored population in the Eastern Health District increased slowly but steadily from 1933 to 1939. The white families moving away were probably not replaced in like numbers, and the average white family remaining was therefore more stable or more permanent than the average colored family in the District.

In grouping families according to *socio-economic class* there seems to be a high degree of association between size of the group and proportion remaining. Skilled workers and foremen are the most numerous and the most permanent. With respect to the lower economic classes, permanence may represent an achievement akin to success in attaining economic stability. For the white-collar workers and professionals, the relative instability of such families may be taken as evidence that the Eastern Health District is not naturally their most appropriate residence as far as the whites are concerned. It is obvious that frequency and permanence are not universally associated. It is always possible to select a group so that the most stable will constitute a small minority. But whatever the specific factors may be which are responsible for migration, or for permanence, if a given characteristic represents an unnatural or undesirable state, then the natural tendency will be for such families to move away to more congenial surroundings, or for that state to change and perhaps destroy the identity of the family in the process.

Of the various characteristics associated with permanence of residence, *ownership of the home* is one of the strongest factors. For this reason, each of the other factors studied in this paper has been correlated with home ownership (for whites) as well. The difference in permanence between white and colored practically disappears when made specific for ownership of the home. This finding is not at all surprising, and it points to the observation so frequently encountered in comparisons between Negroes and whites, viz., that the difference between the races in many respects is chiefly a socio-economic one. We

might be justified in saying that Negro families are less stable because such a small proportion own their homes. Correlation of course does not necessarily indicate causal relationship. In the case of home ownership and race or color, it is not illogical to assume that a family aims at home ownership as a desirable attainment. In most cases, success in this endeavor may be the direct cause of permanence of residence. Certainly migration, enforced by seeking occupation or more congenial surroundings elsewhere, may be the cause of not purchasing a home.

Age of the family head is also one of the most important characteristics associated with permanence of residence, and one where one is tempted, with some logic, to identify correlation with cause and effect. Young family heads are more likely to change jobs; they are less likely to be home owners; and they are therefore more likely to move. Towards middle age they have larger families, their habits become more settled, they may purchase homes, and thus become more permanent residents. Towards old age the family heads die off, and one would expect the families to be lost from the study more quickly due to breaking up or otherwise losing their identity under the definitions used in this analysis. The age of the head is a characteristic peculiar in itself when compared with other bases for classifying families, such as nativity, education, etc. Since this study is applied to a given cohort of families, and change in age is inevitable and in no way subject to random or uncontrolled factors, the length of the interval studied has a direct effect on the reasons to be ascribed to loss of families from the study. In the present instance, with an average 3-year interval, the proportions remaining among old family heads are considerably less than among young family heads (even considering home owners and home renters separately). Since at the older ages death of the head is an important factor, this cause would be materially decreased if the interval were shorter—one year, let us say. Then the proportions remaining in the older groups might be expected to be as high as, if not higher than, those for the youngest groups.

The differences in proportions remaining when families are grouped according to *nativity,* or according to *education* of the family head seem to be due almost entirely to the factor of home ownership. The foreign-born are the most permanent and also have the highest percentage owning their homes; and the same thing is true of families whose heads received only three or four years of schooling.

Size of family, and *number of wage-earners* in the family both seem to support the principle of "the larger the better," from the point of view of permanence of residence. In neither case is this accounted for entirely up the percentage owning their homes. Assuming that stability represents a desirable state, this might be used as evidence that large families and more than one wage-earner are characteristics associated with family well-being.

* * * * * *

There are many instances in which the investigation of a problem in hygiene or public health requires that families be followed for a given length of time—perhaps a number of years; or that a given survey of a population be repeated after a certain interval for purposes of comparison. When the family is the basic unit of such an investigation, it is important to have an indication of the extent to which families originally studied can be identified at a later date.

Table 7 lists the various family characteristics analysed in this paper with respect to permanence of residence. It shows in a general way the categories with the highest and lowest proportions remaining, regardless of cause or association of such classes with other classes analysed here. In the planning of a follow-up study it may be desirable to omit from the investigation families which appear to be the least stable, provided this is not inconsistent with the purpose of such a study. In order to direct an investigation toward the most permanent families in the population, it is convenient that in certain categories, such as color, family type or socio-economic class, the most stable are also the most numerous classes. Of the categories listed in Table 7 as having the highest "stability," some classes exhibit much higher proportions remaining than others. Thus white home-owner families, either married (Table 4), or with 9 or more members in the family (Appendix, Table 15), or with 4 or more wage-earners (Appendix, Table 16), each showed over .9 as the proportion remaining at the end of 3 years. These families would be the most desirable as the base population for a follow-up study, provided they constitute a sufficiently large proportion of the total families, and do not introduce a serious bias. Of these three categories, however, even the largest—the married white home-owner families—constitute less than 30 per cent of the total families in the district.

TABLE 7

Family characteristics associated with the highest and the lowest proportions remaining

CLASSIFICATION	CATEGORIES	
	Highest proportion remaining	Lowest proportion remaining
Color	White	Colored
Home ownership	Home owned	Home rented
Rental	$25 per month or more	$5 per month or less
Family type:		
Presence of children	With children	Without children
Marital status of the family head	Married	No spouse present
Sex of the family head when no spouse is present	Female	Male
Socio-economic class	Skilled workers and foremen	Professional, or servant class
Age of the family head	50 - 55 years	Under 30, or 70 and over
Nativity of the family head	Foreign-born	Native
Education of the family head	3rd or 4th grade	High school or college
Size of family	9 or more members	1 member
Number of wage-earners in the family	4 or more	None (family on relief)

On the other hand, white home-renters, either with no spouse present (Table 4), or with only one member in the family (Appendix, Table 15), or colored families paying less than $10 per month rent (Appendix, Table 8), or with male heads with no spouse present (Table 4), each showed a proportion remaining at the end of 3 years of less than .4. These families would be especially to be avoided in a follow-up study, if we had the choice.

In any event it is useful to have some knowledge beforehand of the extent to which families in given categories may be lost from the study,

and to be able to foresee what effect this might have on the expected results, with the opportunity to do a certain amount of selecting with respect to relative permanence of residence.

REFERENCES

(1) BADGER, G. F. Department of Biostatistics, School of Hygiene and Public Health, The Johns Hopkins University. Unpublished data.

(2) DENSEN, P. M. Family studies in the Eastern Health District. II. The accuracy of statements of age on census records. *Amer. Jour. Hygiene*, Vol. 32, No. 1, Sec. A, 1-38, July, 1940.

(3) PERROTT, G. ST. J., C. TIBBITTS, and R. H. BRITTEN. The National Health Survey—Scope and Method. *Public Health Reports*, Vol. 54, No. 37, p. 1663, Sept. 15, 1939. (Reprint No. 2098).

(4) PHAIR, J. J. The selection of tuberculosis in the Eastern Health District. Thesis submitted for the degree of Doctor of Public Health, School of Hygiene and Public Health, The Johns Hopkins University, May, 1938.

(5) REED, L. J., W. T. FALES, and G. F. BADGER. Family studies in the Eastern Health District. I. General characteristics of the population. *Amer. Jour. Hygiene*, Vol. 37, No. 1, 37-52, January, 1943.

(6) RIDER, R. V., and G. F. BADGER. Family studies in the Eastern Health District. III. A consideration of issues involved in determining migration rates for families. HUMAN BIOLOGY, Vol. 15, No. 2, May, 1943.

(7) VALAORAS, V. G. A comparison of two groups of urban families which differ as to permanence of their residence. Thesis submitted for the degree of Doctor of Public Health, School of Hygiene and Public Health, The Johns Hopkins University, May, 1936.

APPENDIX

TABLE 8

Distribution of families and proportions remaining at the end of 2.5 years, by color and rental (home renters, 1933 only) (Figs. 2 and 3)

MONTHLY RENTAL ($)	WHITE				COLORED			
	Original families		Proportion remaining	σ	Original families		Proportion remaining	σ
	Number	*Per cent*			*Number*	*Per cent*		
Total (known)	4 336	100.0			2 785	100.0		
0 - 9	157	3.6	.420	.039	348	12.5	.394	.026
10 - 14	795	18.3	.536	.018	1 102	39.6	.566	.015
15 - 19	1 197	27.6	.560	.014	590	21.2	.634	.020
20 - 24	1 131	26.1	.604	.015	635	22.8	.635	.019
25 - 29	601	13.9	.596	.020	110	3.9	.673	.045
30 - 34	235	5.4	.602	.023				
35 +	220	5.1						

TABLE 9

Distribution of families, by color and socio-economic class (1933 + 1936) (Fig. 4)

SOCIO-ECONOMIC CLASS	WHITE		COLORED	
	Number	*Per cent*	*Number*	*Per cent*
Total families	21 767	100.0	5 897	100.0
Professional	660	3.0	77	1.3
Proprietors, officials, etc.	3 047	14.0	183	3.1
Clerical workers, etc.	2 919	13.4	67	1.1
Skilled workers and foremen	6 621	30.4	328	5.6
Semi-skilled workers	5 579	25.6	1 109	18.8
Other laborers	1 471	6.8	2 661	45.1
Servant class	629	2.9	1 320	22.4
Unknown occupation, or none	841	3.9	152	2.6

TABLE 10

Proportions of families remaining at the end of 3 years, by color and socio-economic class (Fig. 5)

COLOR AND SOCIO-ECONOMIC CLASS	PERCENTAGE OWNING THEIR HOMES	ORIGINAL NUMBER OF FAMILIES	PROPORTION REMAINING	σ
White:				
Professional	54.0	660	.620	.019
Proprietors, officials, etc.	70.3	3 047	.762	.008
Clerical workers, etc.	61.0	2 919	.714	.008
Skilled workers and foremen	66.1	6 621	.778	.005
Semi-skilled workers	55.0	5 579	.718	.006
Other laborers	52.0	1 471	.740	.011
Servant class	47.1	629	.646	.019
Colored:				
Professional		77	.714	.052
Proprietors, officials, etc.		183	.612	.036
Clerical workers, etc.		67	.657	.058
Skilled workers and foremen		328	.625	.027
Semi-skilled workers		1 109	.589	.015
Other laborers		2 661	.621	.009
Servant class		1 320	.564	.014

TABLE 11

Proportions of white families remaining at the end of 3 years, by ownership of the home and socio-economic class (Fig. 6)

HOME OWNERSHIP AND SOCIO-ECONOMIC CLASS	ORIGINAL NUMBER OF FAMILIES	PROPORTION REMAINING	σ
Home owned:			
Professional	356	.767	.022
Proprietors, officials, etc.	2 139	.832	.008
Clerical workers, etc.	1 769	.820	.009
Skilled workers and foremen	4 366	.861	.005
Semi-skilled workers	3 061	.832	.007
Other laborers	761	.850	.013
Servant class	296	.804	.023
Home rented:			
Professional	303	.446	.029
Proprietors, officials, etc.	902	.596	.016
Clerical workers, etc.	1 133	.550	.015
Skilled workers and foremen	2 236	.616	.010
Semi-skilled workers	2 508	.578	.010
Other laborers	703	.620	.018
Servant class	332	.503	.027

TABLE 12

Proportions of families remaining at the end of 3 years, by color, home ownership and age of the family head (Figs. 7, 8 and 9)

HOME OWNERSHIP AND AGE OF THE FAMILY HEAD	WHITE					COLORED				
	Percentage owning their homes	Original families		Proportion remaining	σ	Original families		Proportion remaining	σ
		Number	*Per cent*			*Number*	*Per cent*		
Total (known):		21 324	100.0			5 689	100.0		
0 - 29	17.1	2 301	10.8	.526	.010	1 086	19.1	.477	.015
30 - 39	46.9	4 925	23.1	.721	.006	1 970	34.6	.607	.011
40 - 49	66.7	5 760	27.0	.802	.005	1 529	26.9	.649	.012
50 - 59	75.6	4 240	19.9	.804	.006	693	12.2	.681	.018
60 - 69	79.6	2 814	13.2	.758	.008	310	5.4	.636	.027
70 +	79.8	1 284	6.0	.588	.014	101	1.8	.465	.050
Home owned:									
0 - 29		393		.766	.021				
30 - 39		2 305		.862	.007				
40 - 49		3 832		.879	.005				
50 - 59		3 195		.853	.006				
60 - 69		2 234		.806	.008				
70 +		1 023		.634	.015				
Home rented:									
0 - 29		1 906		.476	.011				
30 - 39		2 605		.595	.010				
40 - 49		1 912		.647	.011				
50 - 59		1 033		.650	.015				
60 - 69		571		.573	.021				
70 +		259		.405	.031				

TABLE 13

Proportions of families remaining at the end of 3.5 years, by color, home ownership and education of the head (1936 only) (Figs. 10, 11 and 12)

HOME OWNERSHIP AND EDUCATION OF THE HEAD	WHITE					COLORED			
	Percentage owning their homes	Original families		Proportion remaining	σ	Original families		Proportion remaining	σ
		Number	*Per cent*			*Number*	*Per cent*		
Total (known):		10 697	100.0			2 979	100.0		
No schooling	63.4	339	3.2	.710	.021	149	5.0	.564	.027
1st or 2nd grade		109	1.0			181	6.1		
3rd or 4th grade	65.2	1 075	10.0	.749	.013	780	26.2	.627	.017
5th or 6th grade	63.5	2 740	25.6	.743	.008	1 070	35.9	.590	.015
7th, 8th, or 9th grade	58.7	4 338	40.6	.718	.007	533	17.9	.578	.021
High school or vocational school	50.9	1 629	15.2	.677	.012	228	7.6	.568	.030
College		467	4.4	.634	.022	38	1.3		
Home owned:									
2nd grade or less		283		.784	.024				
3rd or 4th grade		701		.816	.015				
5th or 6th grade		1 737		.825	.009				
7th, 8th or 9th grade		2 538		.821	.008				
High school or more		1 063		.817	.012				
Home rented:									
2nd grade or less		163		.577	.039				
3rd or 4th grade		374		.623	.025				
5th or 6th grade		999		.602	.015				
7th, 8th or 9th grade		1 788		.571	.012				
High school or more		1 027		.513	.016				

TABLE 14

Proportion of households having members who are not related to the head of the family, by color (1933 + 1936)

	WHITE	COLORED
Total households	21 767	5 897
Households with unrelated members	896	1 200
Percentage of the total households	4.1	20.3

TABLE 15

Proportions of families remaining at the end of 3 years, by color, home ownership and family size (Figs. 13, 14 and 15)

HOME OWNERSHIP AND FAMILY SIZE	WHITE					COLORED			
	Percentage owning their homes	Original families		Proportion remaining	σ	Original families		Proportion remaining	σ
		Number	*Per cent*			*Number*	*Per cent*		
Total:		21 767	100.0			5 897	100.0		
1	65.2	2 367	10.9	.517	.010	1 007	17.1	.409	.016
2	57.4	5 402	24.8	.668	.006	1 961	33.2	.587	.011
3	57.9	4 943	22.7	.738	.006	980	16.6	.616	.016
4	63.1	3 970	18.2	.806	.006	665	11.3	.669	.018
5	65.0	2 372	10.9	.830	.008	494	8.4	.702	.021
6	65.1	1 310	6.0	.850	.010	298	5.0	.735	.026
7	66.2	678	3.1	.851	.014	222	3.8	.748	.029
8	61.0	384	1.8	.828	.019	128	2.2	.734	.039
9 +	57.1	341	1.6	.859	.019	142	2.4	.754	.036

TABLE 15—*(Continued)*

HOME OWNERSHIP AND FAMILY SIZE	WHITE					COLORED			
	Percentage owning their homes	Original families		Proportion remaining	σ	Original families		Proportion remaining	σ
		Number	*Per cent*			*Number*	*Per cent*		
Home owned:									
1		1 540		.599	.012				
2		3 087		.789	.007				
3		2 857		.855	.007				
4		2 501		.896	.006				
5		1 536		.915	.007				
6		850		.920	.009				
7		449		.909	.014				
8		232		.914	.018				
9+		194		.938	.017				
Home rented:									
1		821		.360	.017				
2		2 295		.504	.010				
3		2 075		.576	.011				
4		1 460		.651	.012				
5		829		.674	.016				
6		456		.722	.021				
7		229		.738	.029				
8		148		.703	.038				
9+		146		.753	.036				

TABLE 16

Proportions of families remaining at the end of 3 years, by color, home ownership and number of wage-earners in the family (Figs. 16, 17 and 18)

HOME OWNERSHIP AND WAGE-EARNERS PER FAMILY	WHITE					COLORED				
	Percentage owning their homes	Original families		Proportion remaining	σ	Original families		Proportion remaining	σ
		Number	*Per cent*			*Number*	*Per cent*		
Totals:		21 732	100.0			5 889	100.0		
0	63.4	3 665	16.9	.641	.008	1 186	20.1	.493	.015
1	56.3	12 904	59.4	.726	.004	3 536	60.1	.615	.008
2	68.4	3 659	16.8	.796	.007	1 008	17.1	.650	.015
3	75.7	1 145	5.3	.861	.010	159	2.7	.799	.032
4 +		359	1.6	.883	.017				
Home owned:									
0		2 321		.710	.009				
1		7 244		.844	.004				
2		2 494		.880	.007				
3		865		.911	.010				
4 +		307		.906	.017				
Home rented:									
0		1 338		.520	.014				
1		5 621		.574	.007				
2		1 151		.614	.014				
3		278		.709	.027				
4 +		52		.750	.060				

COMPLETED GENERATION REPRODUCTION RATES

BY T. J. WOOFTER*
Director of Research
Federal Security Agency

THIS article explores the possibility of a fresh insight into fertility by using completed generation reproduction rates as the measure. It outlines the trend of generation fertility in the United States since 1860, and develops the possibility of generation rates by order of birth and parity of women.

All generally used fertility rates are based on the behavior of a particular group of women at a particular time. The true rate of natural increase is derived from the birth and death rates and sex ratio at birth of a single calendar year. Rates which are based on the ratios of children under 5 years of age to women measure the fertility of the group for the five years prior to the enumeration as reduced by mortality in the ages 0 to 4 during that time. The most widely used gross reproduction rate (which, for purposes of distinction, we shall refer to in this article as the calendar year gross reproduction rate) is based on the age-specific fertility rates of a single calendar year. Thus, the calendar year gross rate for 1940 may be defined as the sum of the 1940 age-specific birth rates of female infants per 100 women of each single year of age in 1940. (It is also sometimes expressed in terms of rates per 1,000 women or per woman.) Computation is more frequently carried out in terms of the fertility rates of 5-year age groups of women, in which case the sum of these rates is multiplied by 5 in order to allow for the fact that five years elapse while the women are passing through this age group. Thus, the calendar year rate is the total of the female fertility rates of that particular year per 100 women in the childbearing ages, assuming the same number of women in each age. This rate is also

* The author is indebted to R. J. Myers and L. O. Shudde of the Office of the Actuary, Social Security Administration, for helpful suggestions and the calculation of generation survival rates in Table 3A.

often interpreted as the number of daughters which would presumably be born to a cohort of 100 women living through the childbearing period, under the assumption that the age-specific rates of the calendar year are constant throughout the years in which the women are in the childbearing ages. In this sense, the rate of a single calendar year is assumed to forecast the performance of a generation.

The conversion of the calendar year gross rate into a net rate involves the further assumption that the mortality rates of the calendar year under consideration will remain constant throughout the survival span of the generation—i. e., the age-specific fertility rate of each age group of mothers is multiplied by the current survival rate from age 0 up to that age. This yields an estimate of the number of daughters per 100 women passing through the childbearing period who will survive to the age at which their mothers gave them birth or, stated another way, the number of girl babies that 100 girl babies just born will give birth to in their lifetime.

In general, the objective of generation rates is similar, namely, to evolve measures of the gross number of daughters, or the net number, who will survive to the age at which their mothers gave them birth. Fertility and mortality rates of a single calendar year are, however, not held constant. The fertility rates of the mothers are those of the calendar years when they attain successive ages—i. e., in the case of women who were 15 years of age in the 1915 generation, the fertility rate for that age in the year 1915 is used; for women 16 years of age in that generation, the rate for 1916 is used, etc., up to the rate for the women 44 years of age in 1944 (cf. Table 1). The sum of these single year rates for the appropriate calendar years of the childbearing period equals the generation gross reproduction rate, or the total number of daughters born to the cohort of women entering age 15 in 1915 and passing through subsequent ages up to age 45.

The conversion of the generation gross rate to generation net rate discards the assumption of constant mortality rates and applies to each cohort of daughters the mortality rates appropriate to the calendar years of their survival—i. e., the daughters of mothers who were 15 years of age in 1915 must live through the calendar years 1915-1929 before attaining their mother's age. Hence, survival rates for these daughters were selected from tables for these years. Likewise, daughters of moth-

ers 16 years of age in 1916 must live through the calendar years 1916-1931, etc.[1]

The generation net reproduction rate is the rate at which the women of a generation are replaced by their daughters who survive to the ages attained by the mothers when they gave them birth.

Thus, the term "beginning year" as used in this paper refers to the calendar year when the cohort of women attains the age 15, and a generation is usually designated by the date of its beginning year. The term "appropriate calendar year" refers to the calendar year when the women or their daughters attain a specific age. In these terms, the difference between the concept of the calendar year rate and the generation rate may be simply stated: The calendar year rate assumes that the fertility and mortality rates of a single calendar year will remain constant throughout the life of a parent and offspring generation (which is never true); the generation rate, on the other hand, assumes nothing,

---

[1] Generation rates are presented in a Study issued by the Statistique Generale de la France as *Etudes Demographiques No. 1*—"Reproduction Nette en Europe depuis l'Origine des Statistiques de l'Etat Civil." (Cf. *Population Index*, April 1946, pp. 74, 80.) This Study develops two different concepts of the generation net reproduction rate which are equally valid but give slightly different results, owing to the different mortality concept involved. They are: (1) calculation with respect to 100 women at the date of their birth of the number of daughters their survivors will probably bear; and (2) calculation with respect to 100 women age 15 of the number of daughters their survivors will probably bear, excluding all such daughters who are estimated to die before age 15. It will be recognized that these concepts differ slightly from that employed in this article, namely, (3) survival of daughters to exact ages of mothers at the time when their daughters were born.

When constant mortality rates are assumed, as in the calendar year net reproduction rates, these concepts are mathematically equivalent, but when improvement in mortality is allowed for by the generation principle, different results are obtained because, age for age, the survival rates of the daughters are more favorable—i. e., in calculating (1) daughters born per 100 women born, the age-specific fertility rates are multiplied by the survival rates of the *mothers* up to those ages. Thus, the survival rates used in the computation of (1) are those of a generation earlier than those used in computing (3). The calculation of (2) involves survival rates intermediate between (1) and (3).

Formula (3) is preferred in this article because in order to compare the full reproductive potential of the generation of mothers and the generation of daughters, survival up to the age of the mothers should be calculated from rates appropriate to the years through which the *daughters* must live. Otherwise, the superior probability of survival of the second generation is not fully taken into account.

but, when data are available, employs the actual rate of the appropriate calendar year for women who attain a particular age during that year. Various assumptions as to future fertility and mortality trends may be introduced if the attempt is made to estimate rates for generations whose fertility behavior or mortality experience extends into the future. The choice of appropriate rates constitutes the only difference between calendar year and generation methods. Computation procedures are the same.

The reproduction rates currently in use are sometimes referred to as measures of reproduction per generation. Whelpton, however, has pointed out that this is a loose use of the term "generation"[2] because, in the current method of calculating reproduction rates, the so-called "generation" is really a cross-section of all women of childbearing ages during a single calendar year. Thus, the traditional method of computation mixes women from a number of generations, some of whom entered the childbearing age only one year before and some 30 years before. To assume that such a cross-section of experience represents a generation is to assume that such experience will be repeated by a cohort which will actually pass through these ages in successive calendar years. Strictly speaking, therefore, the term "generation" should be defined as a cohort of women beginning about age 15 and actually passing through all ages up to about age 45. For example, women who are 15 years of age in the year 1915 should be followed through to the age 45 in the year 1945.

In order to avoid some of the disadvantages which underlie the present types of reproduction rates, Glass has proposed corrections for duration of marriage. Whelpton has proposed refinements for rates of marriage, sterility, and parity. Both of these methods are useful for standardization of annual birth rates by age and should be valuable in further analysis of current fertility. These refinements, however, require somewhat laborious computations and the use of data not generally accessible for many populations and sub-groups. Furthermore, both methods are still based on the behavior exhibited during a single calendar year, and they assume that the effects of marriage ratios and birth orders on fertility will be the same in the future as they are in the calendar year under consideration.

---

[2] Whelpton, P. K. "Reproduction Rates Adjusted for Age, Parity, Fecundity, and Marriage," *Journal of the American Statistical Association*, December 1946.

What neither refinement allows for and what no amount of refinement of ratios based on a single calendar year can allow for is the fact that childbearing covers a span of about 30 years; and, during such a period of time, the generation is subject to varied social and economic climates, and these environmental factors have the most marked effects on reproduction.

It is pertinent, therefore, to suggest that fresh insight may be gained into the analysis of fertility by tossing as many assumptions as possible into the wastebasket and, as Al Smith used to say, "taking a look at the record." If the births of a generation from age 15 to age 45 have actually been recorded, we do not need to assume anything, since we are observing its actual behavior. By doing this, we are enabled to observe directly the effect of the socio-economic climate on marriage rates and the parity of mothers and to observe further how these factors have actually affected the childbearing record of a generation which has actually lived through the childbearing period.

THE DATA

Fortunately for students in the United States, the record of age-specific birth rates now covers a sufficient period of time to afford fairly accurate observations as to the actual behavior of several generations by following them from the beginning to the end of the period. The native white age-specific birth rate figures published by Thompson and Whelpton in their National Resources Planning Board "Estimates of the Future Population, 1940-2000," page 15, go back as far as 1920. With negligible error, the period may be extended back to 1915 by estimating the birth rate in the ages 15 to 20.

The estimates for the years 1915-1930 have to be further adjusted for incompleteness of the birth registration area in these years when some of the high fertility states were not included. In order to correct for this incomplete coverage, Whelpton prepared a series of estimates published in *Population Statistics—State Data* (National Resources Committee, 1937), showing averages for 1918-21 and 1929-31. In the present article, these estimates and combination of the two were used for 5-year age groups which appear in the generations beginning before 1930. Subsequent figures down to 1941 were taken from "Estimates of the Future Population, 1940-2000," Table B-2. Rates from 1941 to

1946 are from *Vital Statistics Reports* and are corrected for underregistration, and rates after 1946 are estimated. Since the generation beginning in 1920 was age 42 in the year 1947, it is only necessary to estimate their birth rate after age 42 to secure their actual performance also. This range of experience may be extended even farther down to date by estimating future birth rates above age 30. Since less than 30 per cent of the gross fertility occurs above age 30, minor errors in estimates of this portion of the total births will not greatly affect the size of the gross reproduction rate if estimates above this age adhere reasonably well to past experience. In Table 2 the estimated figures are printed in italics. These estimates adhere reasonably closely to the high fertility assumptions of Thompson and Whelpton ("Estimates of the Future Population of the United States, 1940-2000"), except that in the 1930 generation some compensatory increase was allowed because of the low fertility in the early ages. The high rather than the medium assumptions were followed in the 1935 and 1940 generations on the assumption that high birth rates in the early ages might encourage births of the higher order in the later ages. The use of medium fertility estimates rather than high fertility estimates for these upper ages would, however, make a reduction of only about 1 per cent in the gross reproduction rates of these latter generations.

COMPUTATION OF GENERATION GROSS RATES

There are two possible methods of calculating gross fertility by the generation method, both of which are illustrated for part of the generation beginning 1915 in Table 1. In Column 2 of this table, specific rates for single years of age are estimated. These estimates were computed by multiplying the 5-year age-specific rates in the appropriate calendar years, as shown in Column 4, by factors for estimating the ratio of fertility in a single age to fertility in the whole 5-year span. For example, a specific rate for age 15 in Column 2 is obtained by multiplying 53 (in Column 4) by .13. Such conversion factors can be obtained from single-year age-specific rates which are published at irregular intervals in the *Vital Statistics Reports*. Since these factors are estimates, it appears just as accurate and considerably more convenient to measure the fertility of a 5-calendar-year period by the sum of the five appropriate 5-

TABLE 1

*Age-specific fertility rates for generation beginning 1915**

AGE OF MOTHERS	ESTIMATED SINGLE-YEAR RATE	CALENDAR YEAR	5-YEAR RATE
15	6.9	1915	53
16	19.5	1916	52
17	44.5	1917	52
18	82.5	1918	54
19	106.5	1919	50
5-year total	259.9		261
20	141.9	1920	160
21	158.3	1921	164
22	158.6	1922	154
23	159.8	1923	150
24	162.3	1924	155
5-year total	780.9		783
25	158.3	1925	141
26	147.0	1926	136
27	131.6	1927	134
28	124.2	1928	130
29	114.7	1929	133
5-year total	675.8		674

* For discussion of this method, see Whelpton, P. K., "An Empirical Method of Calculating Future Population," *Journal of the American Statistical Association*, September 1936, and Kuczynski, R. R., *The Measurement of Population Growth*, p. 117.

year age-specific rates, shown in Column 4, instead of by the sum of the estimated single-year rates, as shown in Column 2—i. e., in Column 4, aggregate fertility in ages 15 through 19 is approximated closely by adding annual fertility rates for ages 15-19 during five successive calendar years. Since this method is based on the use of the five appropriate 5-year specific rates, it indexes fluctuations within the five calendar years when the women were within that 5-year age span. It will be

noted that there is only a slight difference between the 5-year sub-totals of these two columns. The method applied in Table 2 is, therefore, that of listing age-specific rates for 5-year groups for the appropriate five consecutive calendar years and summing these rates. The grand total indicates the number of children born per 1,000 women who have passed through the childbearing ages. This total is multiplied by .0485 to arrive at the number of *female* births per *100* women in the generation.

In presenting a table such as Table 2, the reader should be warned that the rates shown in the table do not represent single-year age-specific rates but are chronological arrangements of 5-year rates which, when totaled by groups of five calendar years, approximate with a high degree of accuracy the sum of the five single-year specific rates; or, when totaled for all the years in the generation, approximate the grand total of all the single-year specific rates for ages 15 to 44.

THE RESULTS

Accordingly, Table 2 presents for native white women the age-specific birth rates and generation gross reproduction rates for generations beginning 1915, 1920, 1925, 1930, 1935, and 1940. The first three generations are based almost entirely on actual experience; the next two involve some degree of projection of past experience in estimating future fertility above age 30; the 1940 generation rates are frankly based, for the most part, on estimates which were made on the assumption that the course of fertility of this generation in ages 25 through 44 will follow (at a lower level) the pattern of the years after World War I, as exhibited by the generation beginning in 1915. By rearrangement of the figures in Table 2, it is possible to calculate rates for any generation in the period covered. For instance, by dropping the first figure of each 5-year group in the 1915 generation and adding the first figure from each 5-year group of the 1920 generation, the rearranged rates are appropriate to the 1916 generation.

Table 3 compares these full generation rates with the rates applying to the calendar year in which the generation entered childbearing age. This table indicates that instead of a rate of 150 which characterized the calendar year 1915, the full chronological experience of the cohort beginning that year resulted in the rate of 119. From the 1915 to the 1920 generation, there is a significant decrease in the completed fertility from 119 to 109, about 8 per cent, but during the period from

TABLE 2

Fertility rates per 1,000 women by age, total children born and gross reproduction rates for generations beginning 1915 to 1940, native white women, total United States

(Figures in italics are estimates of fertility after 1947.)

MOTHER'S AGE	GENERATION YEAR*	SPAN OF GENERATION 1915-1944	1920-1949	1925-1954	1930-1959	1935-1964	1940-1969
15 - 19	1	53	51	49	48	43	45
	2	52	55	47	45	43	47
	3	52	50	48	44	45	51
	4	54	50	46	41	46	52
	5	50	53	44	43	45	46
20 - 24	6	160	141	134	123	132	133
	7	164	137	128	124	142	147
	8	154	137	124	127	160	140
	9	150	135	116	132	158	*130*
	10	155	135	122	127	146	*130*
25 - 29	11	141	127	115	124	136	*131*
	12	136	122	114	132	155	*131*
	13	134	118	116	146	145	*131*
	14	130	113	120	153	*135*	*131*
	15	133	117	119	140	*130*	*131*
30 - 34	16	94	78	83	100	*80*	*79*
	17	88	77	86	105	*80*	*79*
	18	84	78	91	95	*80*	*79*
	19	79	80	100	*87*	*80*	*79*
	20	81	80	99	*87*	*80*	*79*
35 - 39	21	49	45	56	*47*	*46*	*45*
	22	47	47	60	*47*	*46*	*45*
	23	45	46	55	*47*	*46*	*45*
	24	45	52	*45*	*47*	*46*	*45*
	25	44	54	*45*	*47*	*46*	*45*
40 - 44	26	15	16	*15*	*15*	*15*	*15*
	27	14	16	*15*	*15*	*15*	*15*
	28	14	16	*15*	*15*	*15*	*15*
	29	15	*14*	*15*	*15*	*15*	*15*
	30	16	*14*	*15*	*15*	*15*	*15*
Total children per 1,000 women		2,448	2,254	2,237	2,333	2,366	2,271
Female children per 100 women		118.7	109.3	108.5	113.2	114.7	110.1

* The appropriate calendar year of the birth rate at any age is obtained by adding the generation year minus one to the beginning calendar year of the generation. Thus, the appropriate year for mothers age 25 in the 1930 generation is 1930 plus 11 minus 1, or 1940.

TABLE 3

Calendar year and generation reproduction rates, 1915-1940, native white

BEGINNING YEAR	CALENDAR YEAR GROSS REPRODUCTION RATE FOR BEGINNING YEAR	GENERATION	
		Gross rate	Net rate
1915	150	119	107
1916	149	116	105
1917	149	114	104
1918	148	112	102
1919	146	111	101
1920	147	109	101
1921	149	109	101
1922	140	109	101
1923	140	109	101
1924	141	109	101
1925	138	109	101
1926	132	109	102
1927	128	110	103
1928	126	110	104
1929	119	112	106
1930	118	113	107
1931	113	113	107
1932	109	113	107
1933	103	115	108
1934	106	115	109
1935	105	115	109
1936	103	114	109
1937	104	113	108
1938	107	113	108
1939	105	111	106
1940	108	110	105
1945	117	—	—
1946	131	—	—

1920 to 1940, while the calendar year gross reproduction rates showed an extreme drop from 147 to 108, the completed rates for generations beginning in this period showed a remarkably level trend, only fluctuating by a few points. Table 3 compares the fluctuations in the two series. An outstanding characteristic of the generation rates is their relative stability.

If predictions as to future fertility had been based on the calendar year gross reproduction rate of 105 in 1935, the outlook would have been very gloomy. On the other hand, the generation beginning reproduction in 1935 will probably have one of the highest completed rates in recent years (gross rate 115, net rate 109), because the women 15 years of age in 1935 were from 20 to 24 during the war boom period of 1940 to 1944 and were 25 to 29 in the post-war period 1945-1949. Therefore, they show high specific birth rates in the ages 20 to 30 where the maximum fertility occurs.

Examination of the internal fertility pattern of some of the completed generations in Table 2 will indicate that it so happens that each one has been subject to partially offsetting socio-economic influences at different phases of the childbearing span and, likewise, probably to the tendency for abnormally low or high parities at the beginning of the generation, to be, in part, compensated in the older age groups. For example, the 1915 generation reached the high fertility ages (20 to 29) in the high fertility climate of the post-World War I period, but fertility in the older ages of this generation was reduced during the depression period. There was less compensation in the 1920 generation, since its birth frequencies were lower all the way along the line than 1915. In the generation beginning 1925 and especially in that beginning 1930, the normally heavy fertility in the lower ages was reduced by the coincidence of these ages with the depression years, but this early reduction was largely offset by increases in the older ages during the war and post-war boom.

COMPUTATION OF NET RATES

Inasmuch as these age-specific birth rates have been compiled on a generation principle, it is possible to convert them into net rates by the use of generation life tables (cf. Table 3A). The present method of conversion of gross rates to net rates assumes that the life table of the calendar year under consideration will apply in the future to the daughters who are born during that year. Application of the generation life table enables us to discard this assumption and take into consideration the improvement in mortality rates with the passage of time. Owing to the progressive improvement in mortality, the trend in generation net reproduction rates is even more favorable than the trend in gross rates.

TABLE 3A

1915 generation survival rates ($_xp_0$) for 5-year age groups, 5 calendar-year periods—for calculation of daughters' survival to age of mothers when births occurred. Death rates from U. S. Life Tables to 1940, from Thompson and Whelpton's medium mortality estimates thereafter

CHILD'S AGE[b] x	DATE OF BIRTH[a]					
	1917½	1922½	1927½	1932½	1937½	1942½
0	.896	.916	.928	.939	.949	.957
5	.989	.991	.993	.995	.996	.997
10	.994	.995	.996	.997	.997	.997
15		.994	.995	.995	.996	.996
20			.994	.995	.995	.995
25				.993	.994	.994
30					.993	.993
35						.991
Additional 2.5 years[c]	.997	.997	.997	.996	.996	.995
Survival 0 to age of mother at birth[d]	.878	.895	.905	.911	.917	.916
Total of 5-year female birth rates[e]	12.66	37.98	32.69	20.66	11.16	3.59
Net birth frequency[f]	11.1	34.0	29.6	18.8	10.2	3.3

[a] The date of birth is the central date of a 5-year period—i. e., the female birth rate entered on the next to the last line of the table under column "1917½" is calculated as though all births in the years 1915-19 had occurred on July 1, 1917.

[b] The child's age is the beginning age of the survival period. Calendar year of the appropriate survival rate is the date of birth plus child's age—i. e., the rate .989 (second figure in first column) indicates that the probability was .989 that a child attaining age 5 in July 1922 would live to age 10 in July 1927.

[c] Added to bring children to exact age of mothers at the midpoint of the last 5-year span.

[d] Obtained by successive multiplication of all rates above.

[e] Total births per 100 women in the 5 years (sums of the 5-year groups in Table 2 x .0485).

[f] Product of the 2 numbers above. The sum of frequencies in this line (107.0) is the net reproduction rate.

Net reproduction rates in Table 3 were computed by the short method of multiplying the gross reproduction rate by the survival rate to age 28 during the 18.5th year after the beginning of the generation. The exact conversion of generation gross rates to net rates involves the use of survival factors from a number of different life tables. because the survival span of daughters begins with the first year of the childbearing period of mothers of daughters who will survive to age 15 and extends to 45 years after the date when mothers 45 years of age have their children.

The computation may be considerably shortened by the method suggested, namely, that of measuring the mean survival rate of offspring from age 0 up to ages 15 to 44 by the single survival rate to age 28 from a single life table characteristic of the period when the mean mortality occurs (18.5 years from the generation beginning date). Instead of calculating the survival of each cohort of daughters up to the age of their mothers, it is only necessary to calculate the survival of all daughters for a period equivalent to the mean specific age of mothers.[3] The problem then resolves itself into the determination of the appropriate calendar year in which to measure the mean survival rate of daughters beginning at age 0 when mothers are age 28 and living for 28 years thereafter. This period is characterized as follows: Mothers are age 28, 12.5 years after the time when childbearing begins at age 15.5 (beginning date of the generation). From recent life tables, it may be determined that the age of mean mortality of women who die before age 28 is approximately 6 years. Adding the 12.5-year lag after the date of the beginning of the generation to the 6 years which elapse before the date of mean mortality gives 18.5 years after the beginning of

[3] The age of mean mortality of daughters can be calculated from Table 3A by multiplying each fertility rate by each of the survival rates in the appropriate column. The difference between survivals in the successive 5-year periods gives the number of deaths. The average of these numbers weighted by the ages at which deaths occur is the age of man mortality. The age of mean fertility of mothers can be calculated from Table 2 by weighting the mother's age by the age-specific birth rate at that age. In recent white generations, the mean specific age is about 28; in recent Negro generations, about 26.5. Since this mean fluctuates slightly with the changes in distribution of the ages of mothers, the application of this method requires re-calculation of this mean from time to time and a different calculation for different classes of the population. Or the date of mean mortality may be calculated directly from Table 3A without reference to mothers' mean specific age by weighting the deaths with the dates instead of the ages.

the generation as the appropriate calendar year for survival rates to age 28 which are used to convert the generation gross rates to generation net rates. This short method of converting generation gross fertility rates to net rates in a span of years included in Table 3 yielded results within 1 per cent of the exact net rate as calculated by the long method for several generations.[4] Table 3A illustrates the life table and the method of calculation of the generation net reproduction rate for the 1915 generation. The total of the bottom row of figures in this table gives a net reproduction rate of 107.0 in comparison with 106.9 obtained by the short method.

This method of derivation of the generation reproduction rate indicates that a measure analogous to the female intrinsic rate of natural increase may be quite simply derived: If the M_g designates the mean specific age of mothers and r the annual generation rate of female increase, then the formula is:

$$\log\ (1+r) = \frac{1}{M_g} \log \text{ net reproduction rate, or}$$

$$r = \frac{1}{M_g} \log_e \text{ net reproduction rate.}$$

Since M_g is slightly larger than the mean length of the maternal generation, the resulting annual rate will be slightly smaller than the intrinsic rate of natural increase. This measure is to be interpreted as the average rate of increase in the daughter generation per year in the childbearing span of the mother generation.

The striking fact about the net rates shown in Table 3, besides the relative stability which was previously commented on, is that no native white generation since 1915 shows a net rate below the replacement level. This is in contrast to the low calendar year rates of the 1930's. Only three cohorts, those beginning in 1920, 1921, and 1922, show sufficient effects of the depression to reduce their completed net fertility almost to the level of equilibrium. (The rate for these three years was 101.) From that period on to the generation beginning in 1935 the rise in net fertility was uninterrupted, increasing from 101 to 109. It should be kept

[4] Mortality estimates for dates after 1940 are based on Thompson and Whelpton's medium mortality projections. For method of computation of generation mortality and references, cf. Dublin, Louis I. and Spiegelman, Mortimer, "Current vs. Generation Life Tables," HUMAN BIOLOGY, Vol. 13, No. 4, December 1941.

in mind that these are *native white* rates. Since the foreign-born and the non-white rates are higher, generation net fertility for all women runs above the rates shown in Table 3 (cf. rates for total white women, Table 4). For instance, in the generation of 1922 the native white net fertility reached 101, but total net fertility for all women of that generation was approximately 105. Thus, in every generation from 1915 to 1940, total fertility will be well above the level of population replacement.

A basic factor in the stability of generation net reproduction has been the improvement in mortality. It will be noted from Table 3A that, according to the generation mortality rates, a child born in 1942 would have a probability of survival to age 42 superior to the probability of survival to age 5 of a child born in 1917.

MEASURE OF TOTAL NUMBER OF CHILDREN BY BIRTH ORDER

There is a general relationship between completed generation gross reproduction rates and enumerations of total children ever born to women who have lived through the whole childbearing period, provided the population has not been disturbed by migration. Exact correspondence between measures derived from two such different approaches is not to be expected. The one is the accumulation of births recorded by women who lived during any portion of the previous 30-year period; the other is a report of the number of children ever born to women who have lived up to the date of the enumeration. The relationship between the two would, of course, be disturbed if there is any correlation between the number of children born and future expectation of life of the mother. It would likewise be disturbed by errors in reporting number of children and errors in registration of births. There are certain clues which indicate the possibility that at least in the past there has been a higher death rate among mothers of large families than among mothers of small families. There is, however, no clue as to the extent to which this is associated with clinical causes and the extent to which it is to be accounted for largely by rural-urban differentials and socio-economic differentials in birth and death rates. If, however, the generation which began in 1915 and ended in 1945 is checked against the number of children reported ever born to all white women who were 40 to 44 years of age in 1940, there is agreement between the two figures within about 5 per cent. The Census reported 2,459 children per thousand total white

TABLE 4

*Birth rates per 1,000 total white women, by age and order of birth for generations beginning 1915 and 1930**

MOTHER'S AGE	RATE BY ORDER OF BIRTH PER 1,000 WOMEN							TOTAL RATE
	1st	2nd	3rd	4th	5th	6th or 7th	8th and over	
1915 Generation:								
15 - 19	43.0	9.0	2.0	—	—	—	—	54.0
20 - 24	74.0	50.0	25.0	10.0	3.0	1.0	—	163.0
25 - 29	31.0	38.0	30.0	19.0	12.0	9.5	1.9	141.4
30 - 34	10.2	16.8	16.7	14.2	11.8	16.0	8.4	94.1
35 - 39	3.4	5.2	6.3	6.1	5.6	10.0	12.2	48.8
40 - 44	.8	1.0	1.4	1.5	1.5	2.7	6.0	14.9
Total	162.4	120.0	81.4	50.8	33.9	39.2	28.5	516.2
Total × 5	812.0	600.0	407.0	254.0	169.5	196.0	142.5	2,581.0
Children reported by women 40 - 44 in 1940								2,459.0
1930 Generation:								
15 - 19	36.0	8.4	1.4	.2	—	—	—	46.0
20 - 24	63.0	37.4	17.0	7.1	1.8	.6	.1	127.0
25 - 29	44.5	42.0	29.0	11.0	6.8	5.1	1.6	140.0
30 - 34	12.0	27.0	20.4	12.0	7.7	9.8	5.9	94.8
35 - 39	5.0	10.0	8.6	5.6	4.4	7.0	7.9	48.5
40 - 44	1.1	1.5	1.7	1.6	.2	2.6	5.9	14.6
Total	161.6	126.3	78.1	37.5	20.9	25.1	21.4	470.9
Total × 5	808.0	631.5	390.5	187.5	104.5	125.5	107.0	2,354.5

* From *Vital Statistics, Special Reports,* Vol. 25, No. 9.

women of this group; whereas 2,581 resulted from the registration of births to total white women of this generation[5] (Table 4).

As the age of enumerated women increases, the deficit of children reported is greater. An illustration of this fact may be obtained as follows: Women who were 40 to 44 years of age in 1910 were 70 to 74 years of age in 1940 if they survived. Thus, the two censuses provide enumerations of the number of children ever born, as reported by a group of women who were approaching the end of the childbearing period and by their survivors 30 years later. In the case of the white women in this group in 1910, the number of children reported was 378 per hundred as against only 338 in 1940. There was a deficit of

TABLE 5

*Reproduction rates, total white women, generations beginning 1860-1915 (estimated from number of children ever born)**

BEGINNING YEAR	GENERATION	
	Gross rate	Net rate
1860	272	—
1865	260	—
1870	252	—
1875	233	169
1880	216	157
1885	192	146
1890	182	141
1895	168	134
1900	165	136
1905	154	130
1910	141	123
1915	126	113

* With approximate allowance for under-reporting of children and differential mortality by parity. U. S. Bureau of the Census, *Population—Differential Fertility*, 1940 and 1910.

[5] Some difference is accounted for by the fact that the Census recorded number of children only for married women thus omitting most illegitimate children, many of whom had birth certificates filed.

* From *Vital Statistics, Special Reports*, Vol. 25, No. 9.

some 11 per cent in the number of children reported by the survivors to age 70.

By making approximate adjustments for the errors in reporting and differential mortality by parity in the upper ages, it is possible to estimate completed generation gross reproduction rates for generations beginning in the years 1860 to 1915 (see Table 5). It will be noted from this series that, as measured by the generation rates, the drastic decline in fertility occurred largely before the generation beginning in 1915 in contrast to the leveling off since that date, as shown in Table 2.

It is not only possible to estimate the total number of children born by this generation method, but it is also possible to break this number down by order of birth (Table 4). The generation may be followed through for births of the first, second, third, fourth, and higher orders as well as for total births. Likewise, by cumulated subtraction of these rates from 1,000, it is possible to secure an estimate of the parity of the women in the completed generation. That is to say, the difference between the number of women per thousand reporting a first birth and 1,000 represents the proportion that had no children, and the difference between the number reporting a birth of the first order and the number reporting a birth of the second order is the proportion of women who had one and only one child, etc. These parity estimates are shown in Table 6 for the women in the generations beginning in 1915 and in 1930.[6]

---

[6] The validity of these estimates is affected to some extent by the fact that there is probably some duplication in the reporting of the number of births of the first order on birth certificates. Under certain circumstances, such as the previous birth of an illegitimate child or a birth by a previous marriage, the second child may be reported as the first. This would tend to minimize the estimated number of women who had no children. It would not, however, affect the total number of birth certificates filed or to any appreciable degree the ratios between births of the second, third, and higher orders. Conversely, Census reports on the number of children ever born, since children of single women are not enumerated, omit most illegitimate children, thereby exaggerating the proportion of childless women and reducing the proportion with one child only.

TABLE 6

Percentage of women 45 years of age in 1945 and 1959 who had borne specified number of children and percentage of women 40 to 44 reporting specified numbers in the enumeration of 1940

NUMBER OF CHILDREN	1915-1944 GENERATION*	CENSUS 1940†	1930-1959 GENERATION*
0	18.8	22.5	19.2
1	21.2	18.4	17.7
2	19.3	19.8	24.1
3	15.4	13.6	20.3
4	8.4	9.0	8.3
5 and over	16.9	16.7	10.4
Total	100.0	100.0	100.0

* By successive subtraction of items in lines of Table 4 marked "Total × 5."
† U. S. Bureau of the Census, *Population, Differential Fertility*, 1940 and 1910.

CONCLUSIONS

The fact that the completed generation rates remain at a fairly even level from 1920 to 1940 does not necessarily mean that the fertility trend has remained on an even keel at a gross rate from 110 to 115 all during the period. It will be observed from the course of the calendar year reproduction rates during the period 1915-1946 (Table 3) that this span of years seems to cover parts of two cycles. There may, therefore, possibly be a long-time downward secular trend in the generation rates which is concealed by the effects of the high fertility of the early 1920's and of the middle 1940's.

As a corollary to this conclusion, it can be stated fairly definitely that the period for which such data are available in the United States is too short to warrant dogmatic predictions from the generation data presented herewith. As experience accumulates over the next few decades, however, we should be able to analyze past trends with sufficient clarity to make more accurate projections than are possible from single calendar year rates. No reputable economist would base an analysis

of business cycles on data covering only a part of two cycles. Population analysts should be equally as cautious.

It might be further concluded that the sharp decline in fertility which extended almost uninterruptedly from colonial days to World War I has almost leveled off at a gross rate of about 110 with a possible slight long-time downward trend, but that this rate of possible present decline is neither so drastic nor so alarming as some students in the past have feared and that native white gross rates probably will not reach a level much below 110 for any generation which passed through its young childbearing ages during the war boom period 1940 to 1947 and will complete its fertility experience about 1960 to 1970, unless in the near future changes in the family mores or unusually severe depressions make very abnormal reduction in fertility in the older ages of these generations. It would appear that considering age distribution and socio-economic forces at work, net reproduction reached the minimum of about 101 for the generation beginning 1920 and has risen slightly since.

It may further be stated as a conclusion from the data presented above that commonly used calendar year reproduction rates, although extremely useful for comparing aggregates within any country or for comparing two countries which have passed and are passing through similar socio-economic cycles, are of doubtful validity for long-time comparisons to indicate past trends and especially for future estimates.[7] On the other hand, calendar year rates rather than generation rates usually provide more valid comparisons between populations which have been affected by migration.

One obvious use of such a method when adequate data are available is to analyze the internal structure of the fertility of generations, i. e., to answer such questions as: Do later compensatory tendencies partially offset any abnormally high or low birth rate below the age of 30?

This would lead to the hope that the word "generation" will be discarded in referring to age-specific fertility based on a single calendar year and that the term "generation rate" or "completed generation rate"

---

[7] Myers, Robt. J. "Population, Birth, and Mortality Trends in the U. S." *Transactions of the Actuarial Society of America*, Vol. XLI, Part 1, No. 103, pp. 93-94.

will be reserved for rates of fertility computed from the actual behavior of generations by using age-specific fertility rates for calendar years when those ages were attained.

QUOTATIONS

"DAS GESETZ DER KLEINEN ZAHLEN"*

ONE of the classical illustrations of the Poisson frequency distribution is that of the numbers of men killed by horse kicks in the Prussian Army. The illustration was originally given by von Bortkewitsch in 1898.[1] The original appeared as a pamphlet (*"selbständigen Brochüre"*) and is perhaps not too widely available. At least there is evidence that some writers quote Bortkewitsch at second hand, with occasional unfortunate results. It seems of some interest to quote the example in full, together with enough of a summary of the pamphlet to make Bortkewitsch's ideas clear.

Before considering the horse-kick example, it might be helpful to indicate briefly the problem under consideration and the general outlines of the solution given by Bortkewitsch. I have not followed the

---

* From the Department of Biostatistics (Paper No. 234), School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, Md.

[1] Bortkewitsch, L. von. Das Gesetz der kleinen Zahlen. *Leipzig* (Teubner), 1898.

For the spelling of Bortkewitsch I have followed the title page. In later years the spelling Bortkiewicz is usual (e. g., his "Die Iterationen").

I have found little biographical information; the following is taken from "Wer Ist's," 1928.

"Bortkiewicz, v., Ladislaus, o. Univ.-Prof., Dr. phil.—7. VIII. 68 St. Petersburg.—V: Joseph, Oberst; M: v. Rokicka.—Gymn.; Univ. St. Petersburg, Strassburg, Göttingen, Wien, Leipzig; 95-97 Priv.-Doz. Strassburg; 97-00 Beamt. im russ. Verkehrsmin.; 01-20 ao.; 20 o. Prof. d. Staatswiss. Univ. u. Dir. Staatsw. Stat. Semin. Berlin.—W. Sterblichk. u. Langlebigk. d. orthod. Bevölkg. d. europ. Russl. (in d. Schr. d. Petersb. Akad. d. Wiss.) 90/91; D. mittl. Lebensd. 93; D. Gesetz d. kl. Zahlen 98; Radioakt. Strahlg. als Gegenst. wahrscheinlichkeitstheor. Unters. 13; D. Iterationen 17; D. Bevölkerungswes. 19.—Dt. Dem.—Spez: Nat.-ökon.; math. Statist.—Berlin-Halensee, Johann. Sigism.-Str. 2."

He died in October, 1931.

notation of the original, which would be somewhat unfamiliar to most readers.

He begins, following Poisson, with a derivation of the Poisson frequency distribution, in which

$$p(x) = e^{-\mu}\mu^x/x! \qquad (1)$$

is the probability of observing exactly x events. The expression (1) is reached as a limit from the ordinary binomial distribution. In (1), as is well known, the expected number of events, and the variance of the number of events, are both equal to μ, the parameter of the distribution.

Bortkewitsch then considers the problem of estimating μ from a set of observations. He shows that (assuming all possible values of μ equally likely—a Bayes theorem argument) the most probable value of μ is given by the arithmetic mean of the observations.

He now considers the comparison of the scatter of the actual observations with that predicted by theory. (In this he is following the ideas of Lexis[2] on the binomial distribution.) We have, first, the fact that, as noted above, the expected values of the mean and variance are the same; so we may take as an estimate of the standard deviation

$$s = \sqrt{\bar{x}} \ . \qquad (2)$$

This Bortkewitsch calls the "indirect" estimate. Second, he calls the quantity

$$s' = \sqrt{\Sigma (x - \bar{x})^2/(n-1)} \qquad (3)$$

the direct estimate. He gives the standard errors of these as

$$\sigma_s = 1/2\sqrt{n} \text{ and } \sigma_{s'} = \sqrt{2\mu+1}/2\sqrt{n} \ . \qquad (4)$$

He also points out that it is better, in principle, to consider s^2 and s'^2, of which the standard errors are

$$\sigma_{s^2} = \sqrt{\mu/n} \text{ and } \sigma_{s'^2} = \sqrt{(2\mu^2+\mu/n)} \ . \qquad (5)$$

So far we have considered samples from one and the same Poisson distribution. Suppose now that we have m sets of n observations each, where each set is drawn from a population of the form (1), but where

[2] Lexis, W. Zur Theorie der Massenerscheinungen in der menschlichen Gesellschaft. 1877.

the population parameter μ varies from set to set. Bortkewitsch shows that the direct and indirect estimates of the standard deviation of the whole set of mn values are:

$$\text{Indirect, } s_0 = \sqrt{\bar{x}} \ , \tag{6}$$

$$\text{Direct, } s_0' = \sqrt{\Sigma_{ij} (x_{ij} - \bar{x}_i)^2 / m (n-1)} \ , \tag{7}$$

where $\bar{x} = \Sigma_{ij} x_{ij} / mn$, the general mean, and

$\bar{x}_i = \Sigma_j x_{ij} / n$, the mean of the i-th set.

With this brief summary, we may now consider the horse-kick example and what Bortkewitsch says about it. The example is given on pages 23-25.

"4th Example: Fatalities from horse-kicks in the Prussian Army.

"The following table gives the number of military deaths from horse-kicks, by army corps ("G" indicates the Guard Corps) and calendar years.* [See table, facing page.]

"a) We may proceed in the present case exactly as in the last two. We find:

Yearly Result	Number of cases in which the indicated result happened	was to be expected
0	144	143.1
1	91	92.1
2	32	33.3
3	11	8.9
4	2	2.0
5 or more	—	0.6
	$s_0^2 = 0.70$ (0.05);	$s_0'^2 = 0.73$ (0.09);
	$s_0 = 0.84$ (0.03);	$s_0' = 0.85$ (0.05).

"b) We may, however, discarding the Guard, I, VI, and XI Corps, whose organization differs considerably from the normal,† treat the numbers for the remaining 10 corps as though they related to a single corps; that is, assume

* See "Preussischen Statistik (amtliches Quellenwerk)," vols. 38, 46, 50, 55, 60, 63, 67, 80, 84, 87, 91, 95, 99, 108, 114, 118, 124, 132, 135, and 139.

† The Guard Corps contains, apart from artillery, engineers, and train, 134 infantry companies and 40 cavalry squadrons; the XI Corps has three divisions; the I Corps has 30 and the VI Corps has 25 squadrons, against a norm of 20 squadrons.

	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	Total*
G		2	2	1			1	1		3		2	1			1		1		1	16
I				2		3		2				1	1	1		2		3	1		16
II				2		2			1	1			2	1	1			2			12
III				1	1	1	2		2				1		1	2	1				12
IV		1		1	1	1	1					1					1	1			8
V					2	1			1			1		1	1	1	1	1	1		11
VI			1		2			1	2		1	1	3	1	1	1		3			17
VII		1		1			1		1	1			2			2	1		2		12
VIII		1			1			1					1				1	1		1	7
IX						2	1	1	1		2	1	1		1	2		1			13
X			1	1		1		2		2					2	1	3		1	1	15
XI					2	4		1	3		1	1	1	1	2	1	3	1	3	1	25
XIV	1	1	2	1	1	3		4		1		3	2	1		2	1	1			24
XV		1						1		1	1				2	2					8
Total*	3	5	7	9	10	18	6	14	11	9	5	11	15	6	11	17	12	15	8	4	196

* Marginal totals have been supplied.

that we have a single statistical series of 200 elements, and apply to this the procedure of §4. We obtain:

Yearly Result	Number of cases in which the indicated result happened	was to be expected
0	109	108.7
1	65	66.3
2	22	20.2
3	3	4.1
4	1	0.6
5 or more	—	0.1
	$s^2 = 0.61$ (0.06);	$s'^2 = 0.61$ (0.09);
	$s = 0.78$ (0.04);	$s' = 0.78$ (0.06).

"The agreement of theory with observation leaves nothing to be desired, in case a) as in case b)."

The column of expected numbers in example a) may be misleading. These are *not*, as one might suppose, calculated from a Poisson distribution with a mean of 0.70. Instead, they are obtained by summing the theoretical frequencies for each corps, the latter calculated from the observed totals for each corps. One cannot, therefore, compare observed and expected by the Chi-square test. We may, however, calculate the expected frequencies, assuming that the 14 corps are all alike. We get the following:

Number killed	Observed frequency	Expected frequency
0	144	139.0
1	91	97.3
2	32	34.1
3	11	8.0
4	2	1.4
5+	—	0.2

It is clear that the agreement of observed with expected is satisfactory.

The fact that we have obtained a good fit assuming homogeneity among the corps raises the question as to whether there is any real reason for Bortkewitsch's procedure in a) above. Is there any evidence that the corps do differ in the numbers killed by horse-kicks? We can test this by an examination of the corps totals. It is well known that if quantities are independently distributed in Poisson distributions (with the same or different parameters), their sum also follows the Poisson distribution. We have therefore to inquire whether the fourteen corps totals over the 20 years experience might have come as samples from a single Poisson distribution. The convenient test for this, given by Fisher,[3] is to calculate

$$\chi^2 = \Sigma\,(x - \bar{x})^2/\bar{x}, \tag{8}$$

which will have 13 degrees of freedom. We get

$$\bar{x} = 14,$$
$$\Sigma\,(x - \bar{x})^2 = 382,$$
$$\chi^2 = 27.3,$$

while the 1 per cent point is 27.7. We may conclude, then, that it is likely that there are real differences among the corps.

We can carry this one step further, by examining the totals for the ten corps of normal organization. For these we get

$$\bar{x} = 12.2,$$
$$\Sigma\,(x - \bar{x})^2 = 211.6,$$
$$\chi^2 = 17.3.$$

There are now 9 degrees of freedom; the 5 per cent point is 16.9. There is thus a suspicion that there is still some heterogeneity. Examination indicates that the XIV Corps, with a total of 24 deaths, is largely responsible for the size of χ^2, contributing 11.4 to the total. In the absence of further information, of course, all this remains largely speculation.

---

[3] Fisher, R. A., H. G. Thornton, and W. A. Mackenzie. The accuracy of the plating method of estimating the density of bacterial populations. *Annals of Applied Biology*, ix, 1922. Pp. 325-359.

It is perhaps also of some interest to make a similar examination of the yearly totals. Here there seems to be no reason for excluding any corps. Proceeding as before we get:

$$\bar{x} = 9.8,$$
$$\Sigma (x - \bar{x})^2 = 367.2,$$
$$\chi^2 = 37.5.$$

The 1 per cent level for 19 degrees of freedom is 36.2. It would therefore appear that there are significant year-to-year variations in the number of fatalities for the Prussian Army as a whole.

The real interest of these indications of heterogeneity is not, of course, in the bearing they may have on Prussian hippology, but in the fact that material which apparently, tested as a whole, is homogeneous may on closer examination reveal significant heterogeneity. This is not a new observation, but it is one which will bear occasional repetition.

In justice to Bortkewitsch, it should be added that he did not overlook, in his theoretical treatment, the possibility of heterogeneity of this kind. Indeed, his third chapter, "Die Fehlerexcedenten," is largely devoted to this topic. We may summarize briefly some of his results.

He considers a series of observations,

$$x_1, x_2, \ . \ . \ x_n,$$

each drawn from a Poisson distribution, but with parameters

$$\mu_1, \mu_2, \ . \ . \ \mu_n,$$

not all equal. The variance of the x's is then

$$\sigma^2 = E \{ \Sigma (x - \mu_0)^2/n \} \tag{9}$$

where $E \{ \ \}$ denotes mathematical expectation and μ_0 is the mean of the μ's. It is then easy to show that

$$\sigma^2 = \mu_0 + \sigma_\mu{}^2 \tag{10}$$

where $\sigma_\mu{}^2$ is the variance of the μ's.

It appears from (10) that the effect of variation among the μ's is to increase the variance of the observations. Now suppose, says Bortkewitsch, that we increase the scale of our experience k-fold. All the μ's will be increased k-fold; their variance will be multiplied by k^2; and we shall have

$$\sigma^2 = k\mu_0 + k^2\sigma_\mu{}^2 \tag{11}$$

as the variance of the observations under the new conditions.

If we consider (following Lexis) the quantity

$$Q^2 = \frac{\sigma^2}{\mu_0} = 1 + \frac{\sigma_\mu^2}{\mu_0} = 1 + \lambda^2, \tag{12}$$

a k-fold increase in scale will give

$$Q^2 = 1 + k\,\lambda^2. \tag{13}$$

It is clear from (13) that the larger the scale of our experience (the larger the value of k) the larger will be the value of Q^2, and the more readily can heterogeneity be detected. And, conversely, the smaller the scale of our experience, the more homogeneous it will look. This is the property which Bortkewitsch calls the "Law of Small Numbers." We may quote (pp. 35-36):

> "Under the condition of a restricted field of observation one obtains, as we know, a nearly normal dispersion, that is, almost complete agreement between the standard errors calculated by the direct and indirect methods. The smaller the field of observation, and the more rarely the phenomenon in question, as for instance suicide or accident, occurs in a given community, the better do the statistical results fit the appropriate mathematical formula.
>
> "The hypothesis of a variable probability or variable expectation helps us to recognize this behavior as lawful, and in this sense we may well call the fact, that small numbers of events (out of very large populations) are subject to or tend toward a definite norm of variation, the *Law of Small Numbers.*"

The quotation shows clearly that Bortkewitsch was not, as is sometimes supposed, giving a name to the Poisson distribution. He was giving a name to a phenomenon of statistics which he had, or thought he had, discovered.

CHARLES P. WINSOR